Acquisition du français (L2) chez des adultes à Montréal

Volume 1

Études morphosyntaxiques et discursives

N. Bazergui, K. Connors,
M. Lenoble, B. Majkrak

Office de la langue française

Québec

ERRATA

Volume 1

- p. 9 : 21e ligne,
«...**Sankoff-Cedergren**, archiconnu...»

23e ligne,
«...transcriptions d'un **sous-ensemble**...»

- p. 11 : 2e paragraphe, 8e ligne,
«...selon un protocole élaboré par L. **Jasmin-Demers**...»

- p. 14 : dernier paragraphe, avant-dernière ligne,
«...pour ce qui est des facteurs **socio-démographiques**...»

- p. 15 : dernier paragraphe, 4e ligne, «...et qui ont été **ré-interviewés** en 1984...»

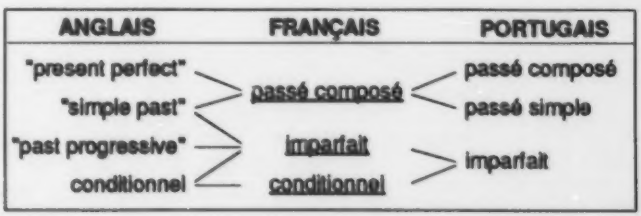
- p. 17 : 1er paragraphe, 3e ligne,
«...combler **quelques-unes** des...»

- p. 26 : 2e paragraphe, 1re ligne,
«Les deux articles complémentaires de N. BAZERGUI **examinent**...»

- p. 29 : 4e paragraphe, 5e ligne,
«Dans la synthèse **ci-dessous**...»

- p. 59 : 2e paragraphe, 1re ligne,
«(4) L'extrant de validation...»

- p. 70 : **TABLE C CORRESPONDANCES CONTEXTUELLES PARTIELLES**



Volume 2

- p. 267 : 1re ligne de la page,
«1..1 $ Comment êtes-vous venu...»

Acquisition du français (L2) chez des adultes à Montréal

Volume 1

Études morphosyntaxiques et discursives

92-01368/

© Gouvernement du Québec, 1990
Dépôt légal - 2e trimestre 1990
Bibliothèque nationale du Québec
Bibliothèque nationale du Canada
ISBN 2-550-20955-9

Acquisition du français (L2) chez des adultes à Montréal

Volume 1

Études morphosyntaxiques et discursives

N. Bazergui, K. Connors,
M. Lenoble, B. Majkrak

Langues et sociétés

Office de la langue française

Québec

Table des matières

Remerciements - K. Connors 7

Introduction - K. Connors 9

Chapitre 1
Transcription, encodage et normalisation de corpus oraux en vue d'une exploitation linguistique semi-automatisée - M. Lenoble 35

Chapitre 2
Usage du système verbal - N. Bazergui 67

Chapitre 3
Acquisition et usage des constructions pronominales - N. Bazergui 81

Chapitre 4
Quelques aspects de la cohésion textuelle : formes pronominales - B. Majkrak 93

Chapitre 5
Quelques aspects de la cohésion textuelle : paraphrase, récurrence, parallélisme, ellipse - B. Majkrak 109

Chapitre 6
Étude corrélative de la variation grammaticale et discursive - K. Connors 121

Chapitre 7
Synthèse - K. Connors 151

Bibliographie générale
Acquisition du français L2 chez les adultes - K. Connors 155

Remerciements

Kathleen Connors

Nous tenons avant tout à remercier les organismes subventionnaires qui ont rendu possible la réalisation et la publication de cette recherche : le Comité d'attribution des fonds internes de recherche de l'Université de Montréal de 1984 à 1986, le programme «Défi» du Ministère de l'Emploi et de l'Immigration du Canada en été 1985, 1986 et 1987 et l'Office de la langue française de 1986 à 1989.

Jean-Pierre Nault, John Reighard et Marco-Antonio Vieira ont collaboré étroitement avec nous pendant les stades initiaux du projet.

Nous avons également profité de l'aide de six étudiants, qui ont porté des jugements globaux sur le français de nos apprenants et de celle de six collègues de l'Université de Montréal : David Sankoff nous a donné des conseils statistiques, Gisèle Painchaud et Alison D'Anglejan nous ont prêté leur corpus transcrit d'entrevues auprès d'apprenants allophones et Pierrette Thibault nous a permis d'utiliser, à des fins comparatives, ceux de vingt francophones faisant partie de l'échantillon Sankoff-Cedergren-Thibault de 1984. Finalement, Nathan Ménard et Mia Toker ont collaboré à la direction du projet antérieur qui a permis la cueillette des corpus «anglophone» et «francophone Balablok» de la présente étude.

Enfin, nous voudrions remercier tous les informateurs qui ont bien voulu se prêter aux entrevues dont les données constituent la base empirique de notre recherche.

Introduction

Kathleen Connors

Ce recueil présente, et représente, le travail de plusieurs chercheurs, depuis plusieurs années. Bien avant que les intérêts de notre groupe de recherche actuel se soient arrêtés sur les thèmes étudiés ici, d'autres groupes ont contribué au présent ouvrage, en constituant des corpus du français parlé de certains groupes de locuteurs qui feront l'objet d'études comparatives ci-dessous. Ainsi, c'était dans le cadre d'un projet de recherche sur le français des diplômés de «l'immersion française» qu'un groupe de recherche en linguistique à l'Université de Montréal a recueilli, en 1980, deux des corpus publiés et exploités ici : ceux des jeunes cégépiens anglophones et francophones qui figureront dans nos études. En 1982-83, un groupe de recherche à la Faculté des Sciences de l'Éducation de la même université a constitué un corpus d'adultes allophones qui avaient suivi des cours intensifs en français, dans le cadre du programme «COFI» (Centre d'orientation et de formation des immigrants, Elbaz 1980). Ces collègues nous ont permis de choisir un sous-ensemble de ces apprenants dont les corpus se prêtent à certaines comparaisons que nous présenterons dans cet ouvrage. En 1984, un groupe de recherche au département d'Anthropologie de la même université a remis à jour le corpus SankoffCedergren, archiconnu en sociolinguistique québécoise. Les responsables de ce corpus nous ont permis d'utiliser les transcriptions d'un sousensemble des locuteurs francophones du nouveau corpus, pour fin de comparaison; il en sera question plus bas. Enfin, en 1984-85, certains membres de l'équipe actuelle, avec d'autres chercheurs, ont élaboré le corpus d'apprenants lusophones du français, corpus qui figure dans nos publications depuis 1985.

Au travail fondamental de cueillette des corpus se sont ajoutés plusieurs niveaux de traitement et d'analyse de ces données : les données allophones et francophones obtenues d'autres groupes de recherche avaient déjà été transcrites par des membres des équipes respectives. Dans le cas de nos corpus anglophones, francophones et lusophones, ce sont des membres d'équipes successives associées à une série de projets sur l'acquisition du français langue seconde, au département de Linguistique à l'Université de Montréal, qui ont transcrit, corrigé et, enfin, entré les données sur ordinateur. L'achèvement de cette étape, finalement, a permis le traitement et l'extraction automatiques de certaines données qui font l'objet des études linguistiques présentées dans ce volume.

Les caractéristiques des informateurs et des corpus

La présente série d'études est axée sur les données d'entrevues semi-dirigées, auprès des trois groupes d'apprenants du français L2, mentionnés ci-haut, auxquels on compare, dans certains articles, le langage de deux groupes de francophones. Il s'agit donc de cinq groupes de locuteurs, au total :

CORPUS À L'ÉTUDE

GROUPE	N	SEXE F	SEXE M	ÂGE	ÉTUDES	CLASSE S-E	L1
anglo	23	17	6	+/-18	12 ans	moyenne	anglais
allo	17	10	7	varié	variées	ouvrière et moyenne	diverses langues asiatiques
luso	20	8	12	20-40	variées	ouvrière et moyenne	portugais
franco Balablok	11	8	3	+/-18	12 ans	moyenne	français
franco Thibault	17	7	10	20-40	variées	ouvrière et moyenne	français

1) Un groupe composé de vingt-trois anglophones diplômés de «l'immersion tardive» en français, qui ont aussi fait l'objet d'une série de recherches dirigées par les Professeurs N. Ménard et K. Connors. Ils habitaient certains quartiers mixtes (francophones-anglophones) de l'ouest de Montréal et ils avaient, au moment de l'entrevue, à peu près dix-huit ans, en moyenne. Ils étaient tous étudiants, en première année du CÉGEP. Lors de leurs études au secondaire, qu'ils avaient terminées quelques mois auparavant, ils avaient suivi, très approximativement, la moitié de leurs cours en français. Les quartiers mixtes (francophones/anglophones) où ces écoles étaient situées, faisaient partie de la banlieue «proche» de Montréal, et se classaient «moyens», sur le plan socio-économique. Étant donné la petitesse de l'échantillon de sujets, nous n'avons pas essayé de distinguer des sous-groupes socio-économiques parmi eux. De la même façon, étant donné qu'il y avait dix-sept filles et six garçons dans ce groupe, nous n'avons pas cherché à établir des rapports entre le comportement linguistique et le sexe. Pour des raisons connexes, nous n'envisageons pas d'amalgamer ces données avec celles des deux autres corpus d'apprenants décrits ci-dessous, non seulement à cause du caractère distinct de chaque groupe de sujets, mais aussi parce que les entretiens qui ont suscité les données constituaient, pour chacun des groupes, une tâche différente.

Dans le cas des anglophones (ci-dessus), le corpus dont nous traiterons ici s'appelle «Balablok», le nom d'un court métrage muet que chacun de nos jeunes informateurs anglophones (et plusieurs groupes témoins dont nous ne parlerons que d'un seul ci-dessous) ont dû raconter en français, individuellement, après l'avoir visionné deux fois de suite. Afin d'obtenir une série de récits et de commentaires comparables, l'intervieweur francophone a posé une série de questions, selon un protocole élaboré par L. JasminDemers et N. Ménard, de l'Université de Montréal. Les intervieweurs ont enregistré ces entretiens et notre groupe de recherche actuel les a transcrits et informatisés. «Balablok», tourné par l'Office national du film du Canada, est un film d'animation qui présente l'histoire d'une guerre entre formes géométriques, les «balles» et les «blocs». Même si le protocole de l'interview, destiné à éliciter la narration et l'interprétation du film, était très détaillé, les intervieweurs ont varié les questions et ont permis aux locuteurs de parler assez librement, pendant à peu près un quart d'heure, au

total. On peut donc considérer ces entrevues comme «semi-dirigées».

2) Le deuxième ensemble d'informateurs étudié ici est composé d'un groupe de dix-sept apprenants du français L2, qui sont locuteurs natifs de l'une des langues de l'Asie du sud-est : le vietnamien, le kmer, le laotien, le chinois ou le coréen. Ces langues, on le sait, ne sont pas toutes apparentées sur le plan génétique. Ces allophones font partie d'un groupe de cent treize immigrants adultes qui ont été interviewés (deux fois, pour la plupart des sujets) sous la direction des Professeurs A. D'Anglejan et G. Painchaud, de l'Université de Montréal (Painchaud, d'Anglejan et Renaud 1984). Les dix-sept apprenants dont nous avons étudié la seconde entrevue remplissent des critères précis : parmi les cinquante-trois informateurs dont le groupe d'Anglejan-Painchaud a transcrit les interviews, ce sous-groupe de dix-sept avait été le seul à subir les deux entrevues prévues et, à la fois, à se classer dans la catégorie supérieure de trois catégories d'apprenants du français L2, au début d'un cours intensif (de 30 semaines et de 900 heures) dans le cadre du programme «COFI», à l'intention d'immigrés récents au Québec.

Ce classement se basait principalement sur le nombre d'années de scolarité et le niveau préalable de connaissance du français. En réalité, nous avons choisi d'étudier ce sous-groupe principalement parce que nos efforts préliminaires visant à dépouiller d'autres sous-ensembles des corpus transcrits n'ont pas porté fruit. Il était très important de pouvoir effectuer, du moins implicitement, l'analyse syntaxique des phrases, des propositions et des syntagmes, pour pouvoir, ensuite, y recueillir les données qui feraient l'objet de certaines évaluations que nous décrivons dans ce volume. Le sous-groupe mentionné ci-dessus nous semblait regrouper les informateurs les plus compréhensibles, en présentant les corpus les plus exploitables dans ce sens. Il comprenait dix femmes et sept hommes, d'âges et de niveaux socio-professionnels variés.

L'entrevue auprès de ces diplômés du cours intensif du français pour immigrants était, elle aussi, semi-dirigée. Elle traitait de la vie et, notamment, de la situation de travail de l'informateur depuis la fin du cours, puisque ce second entretien a eu lieu six à sept mois,

environ, après la fin du programme. Encore une fois, le protocole d'interview était relativement rigide, mais l'intervieweur francophone a varié les questions quelque peu, d'entrevue en entrevue, et il laissait parler ceux qui voulaient bien le faire. Les entretiens enregistrés ont duré à peu près une demi-heure chacun. Les corpus ainsi obtenus ont été transcrits, mais, à la différence des deux autres corpus d'apprenants que nous décrivons ici, le corpus d'Anglejan-Painchaud n'a pas été informatisé. Nous verrons, dans une étude statistique de certaines données quantitatives, que cette limitation nous force à considérer le dénombrement des mots orthographiques comme une mesure approximative de la longueur du corpus dans le cas de ce sous-groupe d'apprenants.

3) Nous exploiterons ici également les corpus de nos vingt informateurs portugais qui ont fait l'objet d'études antérieures (Connors, Reighard et Vieira 1985; Connors, Vieira et Reighard 1986; Connors, Vieira, Reighard et Dillinger 1988). Nos informateurs portugais ont tous immigré au Québec, ou dans un autre pays francophone, à l'âge adulte ou adolescent. Ils étaient tous originaires du Portugal continental ou des Açores, et ils habitaient tous Montréal au moment de l'entrevue. Parmi ces vingt apprenants du français, il y a douze hommes et huit femmes, tous dans la vingtaine ou la trentaine à l'époque de l'entrevue. Même s'il y avait de grandes différences de niveaux de scolarité et même de connaissance du français, nous estimons qu'ils étaient, globalement, le sous-groupe le plus fort en français, parmi ceux que nous étudions ici.

La moitié de ces locuteurs représentait la majorité des Montréalais d'origine portugaise en ce qui a trait à la scolarité : ils n'avaient fait au complet que le premier cycle scolaire au Portugal, se terminant en quatrième année. L'autre moitié avait bénéficié d'une scolarisation plus longue, mais la durée et la nature de la formation des différents informateurs variaient beaucoup. Comme dans le cas des sous-groupes d'apprenants mentionnés ci-dessus, la petitesse de l'échantillon nous incite à ne pas le subdiviser en sous-groupes, en fonction du sexe ou du niveau de scolarité.

Parce que nous avons eu énormément de difficulté à recruter ces informateurs et pour les mettre à l'aise, nous avons choisi un intervieweur non francophone qui parlait français couramment et

correctement. Ceci était dû vraisemblablement au fait que la communauté portugaise avait été, au début des années 1980, une des cibles d'accusations selon lesquelles certains immigrants envoyaient leurs enfants illégalement à l'école anglaise. Quoi qu'il en soit, chacun des vingt lusophones que nous avons fini par recruter a répondu (dans une entrevue semi-dirigée toujours) à des questions portant sur son émigration du Portugal, son immigration au Canada, son adaptation au mode de vie d'ici et ses diverses situations de travail. Nous leur avons demandé soit de raconter un conte folklorique, soit d'expliquer un jeu ou une recette et, finalement, de s'imaginer ce qu'ils feraient s'ils gagnaient 1 000 000 de dollars à la loterie. Nous avons fait transcrire le corpus semi-orthographiquement, principalement par l'intervieweur; suivait une série de vérifications par des francophones. Comme le corpus anglophone mentionné ci-dessus, nous l'avons informatisé récemment. Nous verrons ci-dessous les aspects du traitement des données que l'informatisation a automatisés.

Le corpus lusophone correspond très clairement, à bien des égards, aux données, plus largement connues, des *Gastarbeiter* de langue maternelle romane (voir, par exemple, Klein et Dittmar 1979; Clahsen, Meisel et Pienemann 1983). C'est que de tels locuteurs varient beaucoup pour ce qui est de l'âge, de l'âge lors de l'immigration et des qualifications professionnelles. Dans le cas de nos informateurs portugais, une des causes de la variation sur ces points est le fait que certains d'entre eux soient arrivés à Montréal, avec leurs parents, pendant l'adolescence et aient brièvement fréquenté l'école québécoise. Nous traitons de certains de ces corrélats de la variation socio-démographique (et finalement linguistique) ailleurs (Connors et coll. 1988).

Ici, nous voudrions seulement souligner le fait que le groupe anglophone soit uniforme du point de vue de la L1, de l'âge et de la formation scolaire (jusqu'à la méthode d'apprentissage du français), alors que les allophones du corpus Painchaud-D'Anglejan partageaient plutôt le statut d'immigrants récents et de diplômés récents d'un cours intensif de français L2. Les lusophones, en revanche, partageaient essentiellement la nationalité et la L1, comme les anglophones, mais présentaient de grandes variations par ailleurs, pour ce qui était des facteurs sociodémographiques qui sont souvent liés à la variation de la performance linguistique

des apprenants d'une L2 (Klein et Dittmar 1979, Clahsen et coll. 1983, Painchaud et coll. 1984).

4) Les francophones «Balablok» : parmi les nombreux informateurs qui ont subi l'entrevue «Balablok» mentionnée ci-haut, nous avons choisi onze étudiants du même niveau scolaire que celui des anglophones (le groupe 1, ci-dessus). Ce groupe témoin, qui était en première année du CÉGEP dans la concentration «Lettres», étaient représentatifs, à nos yeux, des francophones de l'âge et du niveau scolaire de nos sujets anglophones, tout en représentant la norme du français québécois parlé, dans la mesure où notre protocole d'entrevue suscitait des manifestations de cette norme. Comme nous l'avons indiqué pour les groupes d'apprenants présentés ci-dessus, ce groupe de huit filles et trois garçons, à plus forte raison, ne se prête pas à des subdivisions selon le sexe ou le niveau socio-économique. La transcription et l'informatisation de leur corpus nous ont fourni, globalement, autant d'échantillons de la façon dont les francophones, comparables à nos sujets anglophones, pouvaient s'acquitter des tâches linguistiques exigées par le protocole d'interview «Balablok». La publication, dans cet ouvrage, de leur corpus à côté du corpus correspondant des anglophones, facilite l'appréciation, globale et impressionniste, du niveau de français qu'atteignent les diplômés de l'immersion «tardive». À la différence des deux corpus d'apprenants publiés ici, le corpus francophone «Balablok» ne contient pas les codes numériques qui permettent l'extraction automatique des formes verbales. (Nous expliquerons ces codes dans la présentation des caractéristiques techniques de nos corpus informatisés, ci-dessous). Cette lacune n'est qu'une conséquence des contraintes imposées par les limites de nos ressources matérielles.

5) Les francophones «Thibault» : nous désignons par ce nom un groupe de dix-sept francophones qui constitue un sous-ensemble des informateurs québécois montréalais qui avaient été interviewés en 1971 par le groupe Sankoff-Cedergren et qui ont été ré-interviewés en 1984 par une nouvelle équipe de sociolinguistes du département d'Anthropologie à l'Université de Montréal. (Dans cet ouvrage, ce corpus portera le nom de la coordinatrice du projet de recherche à l'époque de la seconde série d'entrevues, dont nous exploitons les transcriptions fournies par la même collègue.)

Nous avons choisi ces francophones comme groupe témoin, à faire correspondre, dans la mesure du possible, aux apprenants lusophones (le groupe 3 ci-dessus). Les protocoles d'interviews des lusophones et des francophones «Thibault» ne se ressemblent cependant qu'à un niveau très général : dans les deux cas, les intervieweurs posaient des questions sur la vie quotidienne (travail, habitation, enfants), mais le point de départ, dans le cas des lusophones, était l'émigration du Portugal et l'immigration au Canada. Pour des raisons évidentes, de tels événements n'ont pas figuré dans les conversations avec les francophones québécois (voir les remarques sur le protocole d'interview à la section 3 ci-dessus).

C'est surtout dans le choix d'un sous-ensemble des informateurs du corpus Thibault que nous avons essayé d'établir une correspondance : l'intervieweur principal des lusophones a trouvé, parmi les francophones en question, un sous-groupe qui correspondait à ceux-là pour ce qui est de l'âge approximatif, la répartition des sexes et les niveaux socio-économiques généraux (classe ouvrière ou classe moyenne, les deux étant représentées chez les Portugais de notre échantillon). Il était impossible d'apparier les informateurs lusophones et francophones en fonction du niveau de scolarité, les systèmes scolaires étant trop peu semblables, surtout en ce qui concerne l'âge auquel les ressortissants des classes défavorisées quittaient l'école à l'époque pertinente (voir la section 3 ci-dessus). L'appariement selon l'âge au moment de l'interview, que nous avons opéré dans ce cas, mise sur le postulat sousjacent selon lequel les apprenants prennent inconsciemment comme modèle des membres de leur groupe d'âge dans la population cible. De toute façon, l'usage que nous avons fait du corpus «francophone Thibault» n'implique que des traits linguistiques ayant, normalement, une distribution très générale : il ne s'agit pas d'indices d'appartenance à des sous-groupes sociolinguistiques très restreints. (Voir surtout l'article de Bazergui sur les verbes «pronominaux», ci-dessous).

L'état de la question

Dans le premier volume de ce recueil, nous présentons un ensemble d'études sur l'acquisition du français langue seconde par des Montréalais adultes ayant diverses langues maternelles. Même

si notre recherche partage les préoccupations principales du domaine de l'acquisition des L2 en général, nous avons voulu, tout particulièrement, combler quelquesunes des nombreuses lacunes dans nos connaissances sur l'acquisition du français chez les adultes. Le foyer d'intérêt, pour nous, est la nature des problèmes inhérents à la caractérisation linguistique des niveaux d'acquisition du français L2. Quelle est la nature des écarts entre la grammaire française des francophones et celle des non francophones? Quels écarts caractéristiques des grammaires «non natives» du français se prêtent à des études quantitatives et qualitatives qui détermineront et caractériseront des niveaux d'acquisition de cette langue?

Les écrits sur l'acquisition du français L2 en Amérique du Nord se limitent, le plus souvent, à des évaluations globales (ex. les appréciations du français des élèves canadiens qui suivent des programmes «d'immersion française» Holobow et coll. 1987, Genesee 1987, ou, plus rarement, des immigrants : D'Anglejan et coll. 1985, 1986). Quelques travaux, cependant, ici et en Europe, décrivent des faiblesses ou des particularités du français d'apprenants (adultes ou enfants), sans nécessairement prétendre découvrir des traits diagnostiques des niveaux d'acquisition (Singleton 1987, White 1987, Trévise et Porquier 1986, Piétropaolo-Saura et Roffé 1985, Trévise et Noyau 1984, Véronique 1984). En nous inspirant de certaines recherches récentes sur l'acquisition de l'allemand (surtout Clahsen et coll. 1983), nous avons voulu explorer l'hypothèse selon laquelle certains traits ou structures linguistiques de distribution générale seraient des indices critiques de stades ou de niveaux d'acquisition. Une telle perspective (à condition de ne pas faire fausse route) permet de combiner des analyses linguistiques fines avec la description générale (et éventuellement utile) des grammaires des apprenants.

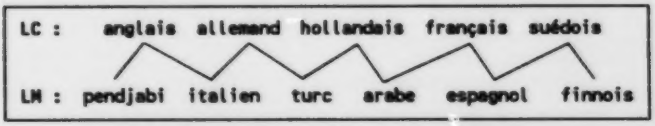
La préoccupation principale de la phase actuelle de notre programme de recherche, dont nous voudrions faire le compte rendu ici, est une certaine intégration de la récente perspective «communicative» à l'évaluation de la performance du français langue seconde. De plus en plus répandues, depuis dix ans, dans le milieu de la pédagogie des langues étrangères, les évaluations de la compétence «communicative» se distinguent des jugements portant sur la grammaticalité. S'appuyant sur l'observation quasi-

universelle de la dissociation entre la connaissance consciente des règles d'une langue étrangère et la performance efficace dans cette langue, l'étude de la compétence à communiquer en langue seconde se veut une évaluation de l'efficacité relative, à tous points de vue, de la performance linguistique.

Or, une telle compétence a nécessairement une composante linguistique, et plus particulièrement, un contenu spécifique à la langue seconde en cause. Autrement, l'évaluation de la compétence à communiquer n'aurait pas sa place dans l'étude de l'acquisition des langues secondes. Ce qui nous intéresse ici, c'est la nature de cette composante linguistique et le rapport entre les capacités sous-tendant la grammaticalité, d'un côté, et la «communicativité», de l'autre. En d'autres termes, quels sont les liens (puisque nous présumons qu'il y en a) entre la compétence relative à construire des phrases grammaticales en L2 et la capacité relative de construire des énoncés efficaces (au sens large) dans la même langue? Quels sont les fondements linguistiques, en L2 spécifiquement, de ces deux types de compétence, que l'observation ordinaire juge comme dissociables et souvent dissociés? Quelle est la nature des rapports qui doivent, logiquement, exister entre eux?

Nos études s'apparentent, pour ce qui est de leur contenu, à certaines recherches récentes en Europe : sur le plan des objectifs très généraux, elles ressemblent en partie à celles découlant du récent projet subventionné par la "European Science Foundation" (dorénavant ESF, Perdue 1984, 1987). Il s'agit d'un ensemble d'études sur l'acquisition de certaines L2 (le néerlandais, l'anglais, le français, l'allemand et le suédois) par des travailleurs migrants ayant certaines L1, (l'arabe, le finnois, l'italien, le pendjabi, l'espagnol ou le turc). Plus précisément, pour chaque L2, les chercheurs étudiaient son acquisition par des locuteurs de deux L1, et pour quatre des L1, on examinait l'acquisition de deux des L2. C'est-à-dire qu'en Allemagne, des linguistes germanophones comparaient l'acquisition de l'allemand par des locuteurs du turc et de l'italien; en Angleterre, il s'agissait de l'italien et du pendjabi; en France, des linguistes francophones comparaient l'acquisition du français par des arabophones et des hispanophones, alors qu'aux Pays-Bas, on examinait l'acquisition du néerlandais par des Turcs et des arabophones et finalement, en Suède, l'acquisition du suédois par des Espagnols et des Finlandais.

Ainsi, pour chaque L2, on étudiait des apprenants ayant comme L1 une langue étroitement apparentée à la L2 et des apprenants ayant une L1 moins apparentée, ou d'une toute autre famille. À l'inverse, pour quatre L1, on étudiait deux groupes de sujets parlants, qui apprenaient deux L2 différentes. Ainsi, les chercheurs du projet ESF représentent le choix des langues en présence de la façon suivante (voir, p. ex. : Deulofeu et Noyau 1986 : 5) :



LC = Langue cible LM = Langue maternelle

Pour chaque «paire» de L1-L2 (p. ex. arabe-français), on a suivi le développement en L2 de quatre apprenants pendant deux ans et demi. Les enregistrements correspondant à cette étude longitudinale constituent le corpus principal de l'équipe ESF.

Dans la mesure où nous nous livrons à des études quantitatives et qualitatives sur des structures et des éléments spécifiques au français, nos recherches correspondent de plus près à certaines études associées au GRAL (Groupe de Recherches sur l'Acquisition des Langues, dont les membres ont participé au projet ESF). Ces recherches sont exposées dans plusieurs revues et quelques recueils récents : Trévise 1982, Andersen 1984, Giacomi et Véronique 1986, *Langue française* 71 (1986), *Langages* 84 (1986), *Studies in Second Language Acquisition* 8 : 3 (1986). Pour leurs travaux antérieurs au projet ESF, voir Perdue et Porquier (1979, 1980) et Trévise (1982).

Même si certains chercheurs du côté européen s'intéressent, comme nous, aux jeunes adultes qui acquièrent le français L2 aux études également (Raupach 1984, Möhle 1984, Py 1984, Grandcolas et Soulé-Susbielles 1986, Towell 1987), la recherche en Europe oppose, généralement, l'acquisition «dirigée» à l'acquisition «spontanée» (ou «guidée» vs «naturelle» - Deulofeu et Noyau [1986]). C'est-à-dire que, sans nier l'influence provenant des ressemblances ou des dissemblances génétiques et typologiques entre L1 et L2 (Perdue 1984), on postule (avec raison, à notre

avis) que le processus d'acquisition est influencé par des procédés métalinguistiques et par l'enseignement en particulier (le cas échéant). À la différence de certaines études récentes sur l'acquisition de l'espagnol, de l'allemand ou de l'anglais (Giacobbe et Lucas 1982, Berthoud 1982, Pienemann 1987, Lightbown, Spada et coll. 1988), cependant, les recherches publiées sur le français L2 en Europe, tout en insistant sur cette distinction, n'entrent généralement pas dans le détail des rapports entre les structures enseignées et la performance linguistique des élèves. Il s'agit simplement de traiter le français des deux types d'acquérants comme des phénomènes distincts. Ce désintérêt relatif pour la nature du français appris à l'école reflète peut-être la réorientation de la linguistique appliquée à la pédagogie en France depuis plus de dix ans : *"The language system lost its priority and became one of many other components in the communication process"* (Grandcolas et SouléSusbielles 1986 : 293).

Ici, au contraire, nous avons implicitement considéré la L1 comme la distinction principale entre nos groupes d'apprenants du français, surtout parce qu'il y a, dans notre cas, un seul sous-groupe qui est partagé entre apprenants «dirigés» et «spontanés». C'est-à-dire que la grande majorité de nos informateurs ont étudié le français (voir description de nos corpus ci-dessus).

La préoccupation principale des études portant sur la performance linguistique d'apprenants d'une L2 est le statut de la variation dans les grammaires de tels locuteurs. S'agit-il de la variation systématique ou aléatoire? Si la variation est systématique, il faut pouvoir en rendre compte dans le détail; il faut pouvoir découvrir et décrire le système dont elle n'est qu'une manifestation. En définitive, il faudra déterminer chaque contexte qui déclenche la variation, en motivant les variantes déclenchées. Dans la mesure où, au contraire, on qualifie la variation d'aléatoire, on admet notre incapacité de l'analyser. Ce qui semble faire l'unanimité parmi les chercheurs, c'est l'importance relativement grande de la variation apparemment aléatoire chez les apprenants d'une L2, due (on le présume) à leur incapacité de simuler les modèles de variation caractéristiques des locuteurs natifs et de l'associer à des contraintes sociostylistiques, qui limitent et canalisent la variation chez ces derniers.

L'existence de la variation aléatoire, sans parler de son importance, pose un problème capital à l'hypothèse classique de l'«interlangue» systématique. Cette hypothèse, très courante pendant les années 1970, attribuait implicitement aux grammaires des apprenants le type de systématicité qu'on trouve chez les locuteurs natifs. Comme Ellis (1986 : 10) l'a bien dit, si la variation est *inhérente* à l'interlangue (c'est-à-dire une variation distincte de celle des locuteurs natifs), elle remet en cause le caractère «naturel» et uniforme du processus d'acquisition de la L2, car l'hypothèse de l'interlangue présumait que la systématicité «intraapprenant» correspondait à une certaine uniformité «interapprenant». En effet, certains spécialistes insistent sur le caractère systématique de la production linguistique des apprenants, malgré son instabilité évidente : cette production serait *"systematic, since it can be described by a set of rules, and unstable, since it is in a constant state of evolution"* (Giacobbe et Cammarota 1986 : 328).

Les études européennes sur le français L2 des adultes

En examinant de plus près les études sur l'apprentissage du français qui ont été faites dans le cadre du projet ESF par l'équipe française (le GRAL - voir p. ex. : Perdue et coll. 1986, 1988), à la recherche de la systématicité et de l'uniformité, on a l'impression que sur le plan empirique, tout reste à découvrir, malgré le nombre et l'importance de leurs articles. Ces travaux sont plus riches en hypothèses, voire en spéculations, qu'en vérifications empiriques. Les données linguistiques dans ces écrits sont éparses et les traitements quantitatifs de données quantifiables se font encore plus rares. C'est une des raisons pour lesquelles nous avons voulu présenter ici des études empiriques dont les résultats seraient directement vérifiables (du moins pour les corpus que nous publions ici). Ce faisant, nous courons le risque d'attirer des critiques certaines : nous aurons sans doute compté les mauvaises formes linguistiques, de la mauvaise manière. Il nous semble, pourtant, que des études quantitatives de certaines formes linguistiques produites par les apprenants dans la langue cible constituent une étape nécessaire à la caractérisation des structures qui sont apprises ou qui, justement, ne sont pas apprises.

En d'autres termes, nous supposons que la description de la compétence des apprenants en L2 passe par l'étude détaillée des formes linguistiques qu'ils emploient avec une certaine fréquence. En définitive, nous voudrions trouver, dans la comparaison des résultats de telles études détaillées, les caractéristiques linguistiques des apprenants par rapport auxquelles la variation constitue la manifestation de la différenciation des niveaux d'acquisition. Face à cet objectif, l'examen des études de l'équipe ESF suscite, chez nous, des critiques qui servent à éclaircir certains problèmes posés par l'importante variation de la performance des apprenants.

La décision, de la part de l'équipe ESF, de caractériser la structure des énoncés de leurs apprenants «globalement» (voir Klein et Perdue 1988) a mené à un problème insurmontable : étant donné la variation énorme à cet égard, d'apprenant à apprenant, les auteurs se sont trouvés dans la nécessité de décrire des grammaires individuelles, basées sur très peu de données (apparemment, selon ce qu'on peut voir ou apprendre de ces données dans les articles pertinents), et tributaires de nombreuses suppositions et interprétations *ad hoc* de la part des analystes. C'est que les apprenants analysés par l'équipe ESF ont généralement atteint des stades ou des niveaux peu avancés de l'acquisition des L2, et/ou ils n'ont pas suivi de cours de langue. De plus, le design expérimental global de la recherche ESF a pour conséquence le fait qu'on n'étudie, en général, que quatre informateurs ayant la même L1 (voir ci-dessus). L'effet produit chez le lecteur est l'impression de lire des «règles» qui ne se vérifieront peut-être plus, au-delà de l'échantillon des données présentées dans le texte des articles.

Cette impression persiste, à la lecture de la plupart des travaux ESF, en dépit du fait qu'ils renferment des observations importantes et intéressantes sur la structure des énoncés des apprenants et, avant tout, sur les éléments ou les constructions caractéristiques des «locuteurs natifs», qui n'y apparaissent pourtant pas. C'est pourquoi il nous paraît essentiel, dans l'étude des productions linguistiques des apprenants, d'identifier des éléments ou des structures qui puissent servir d'indices de l'avancement relatif de l'acquisition de la L2, c'est-à-dire des traits linguistiques dont la variation des manifestations distingue des stades ou des niveaux. Les candidats à ce statut «d'indice» doivent jouir d'un statut obligatoire dans la grammaire de la langue, ou figurer très souvent

dans les énoncés «natifs» de tous les styles ou de tous les registres. Idéalement, les variantes «natives» de ces différentes formes linguistiques devraient émerger (Clahsen et coll. 1983) ou se stabiliser (Brown 1973) dans la grammaire selon un ordre invariable, ou un ordre dans lequel la variation ne joue qu'un rôle restreint et bien cerné.

En l'absence d'études longitudinales, les indices que nous présentons dans les articles de ce recueil ne semblent avoir rempli que les deux premières exigences parmi les trois que nous voulions vérifier (à savoir : la capacité d'ordonner les apprenants selon leur stade ou niveau d'acquisition, le caractère obligatoire ou fréquent dans les productions «natives» et l'ordre d'émergence ou de stabilisation invariable).

Par contre, l'équipe ESF, en choisissant la stratégie «globale», auprès d'apprenants débutants, se trouvait généralement (et surtout au début de l'étude longitudinale) devant des productions qui consistaient, *grosso modo*, en quelques dizaines de mots différents (types lexicaux), puisés dans la L2 et enchaînés dans un ordre qui n'était clairement imputable ni à L1, ni à L2. Le problème principal est le fait que ces énoncés ne soient pas suffisamment autonomes, sur le plan linguistique, pour admettre des interprétations univoques, et ainsi des gloses et des étiquettes syntaxiques.

Ce serait, bien entendu, trop simpliste de vouloir réduire le problème de l'ordre des mots à une question de SVO, SOV ou VSO, selon le soi-disant ordre «de base» ou l'ordre le plus fréquent en L1, ou en L2. L'obstacle posé par la plupart des données exposées et analysées dans les travaux ESF relève de leur caractère indéterminé sur le plan linguistique. Les traductions de ces données en langue cible, proposées par les intervieweurs et les analystes, dépendent étroitement de la connaissance de ce que l'informateur voulait dire et de ce qu'il aurait dû dire, puisqu'il fallait raconter une histoire, déjà connue de l'équipe des chercheurs. C'est pour cette raison que les gloses proposées sont trop spéculatives, dans l'ensemble, pour permettre, à leur tour, l'étiquetage syntaxique.

Face à cette impasse, les analystes présument que les stades pré-syntaxiques sont caractérisés par des ordonnancements et des élisions d'éléments qui sont déterminés par la structuration fonctionnelle des énoncés. (Les analyses fonctionnelles sont présentées en termes de «topique» et de foyer, ou encore, en termes de *"controller"/"experiencer"* et de ce qu'il fait ou de ce qui lui arrive.) Par la suite, la grammaticalisation graduelle rend la structuration des énoncés compatible, en partie, avec la grammaire de la langue cible.

Il ne faut surtout pas perdre de vue le fait que nous présentions ici une simplification abusive de l'énorme hétérogénéité des données qu'on entrevoit dans ces travaux. Le caractère «asyntaxique» des productions au départ est tout relatif chez les hispanophones acquérant le français, par exemple. Il reste, cependant, que l'analyse des énoncés, qu'elle soit syntaxique ou «fonctionnelle», est tributaire des connaissances préalables des chercheurs et n'est que très incomplètement déterminée par les données proprement linguistiques. En d'autres termes, les productions des apprenants ne sont généralement pas interprétables par des moyens linguistiques et, donc, elles ne sont pas analysables de façon univoque, et donc convaincante.

De telles interprétations et analyses seraient bel et bien nécessaires pour atteindre un des objectifs principaux du projet ESF, celui qui est à l'origine du design expérimental exposé ci-dessus : on voudrait pouvoir faire la part de la contribution de la structure de la L1 et de celle de la L2 au processus d'acquisition de cette dernière. Il est possible que cet objectif soit irréaliste pour un projet qui étudie si peu de représentants de chaque paire L1-L2, généralement à des niveaux de débutants, sans cours de langue et sans limites aux sources d'hétérogénéité. Les données provenant d'hispanophones acquérant le français laissent croire qu'il aurait été salutaire de concentrer les efforts sur des comparaisons, plus étroites sur le plan linguistique, de deux types d'apprenants. Ces derniers pourraient se distinguer par la L1, par un écart très évident de stade ou de niveau, ou encore par la fréquentation (ou non) des cours de la L2 (même si les chercheurs ESF présument que cette dernière variable est sans conséquence importante).

Étant donné l'état actuel de nos connaissances, les données du projet ESF semblent utiles surtout pour nous indiquer, indirectement, les particularités des langues cibles que les apprenants (débutants, du moins) n'acquièrent pas, ou les structures qu'ils ne déploient pas automatiquement.

Présentation des articles

M. LENOBLE débute en se penchant ici sur la description de la cueillette et du traitement de nos corpus, tout particulièrement de nos corpus informatisés et publiés ici : ceux des «lusophones», des «anglophones» et des «francophones Balablok» (voir descriptions des corpus ci-dessus). Tout d'abord, il décrit brièvement nos entrevues, du point de vue de leurs objectifs linguistiques. Ensuite, il traite de certains problèmes associés à l'étape de la transcription manuscrite - problèmes qu'il faut régler en fonction de l'utilisation prévue des données. Une grande partie de l'article expose et justifie les décisions techniques que nous avons prises pour uniformiser la présentation des données informatisées et pour permettre le traitement et l'extraction automatiques de certaines formes. La description des phases d'entrée et de validation des données passe, en partie, par une discussion détaillée du logiciel JEUDEMO, employé ici et ailleurs pour certaines phases du traitement, de la vérification et de l'exploitation des données. L'auteur expose la façon dont ce logiciel utilise les caractéristiques formelles de nos corpus (tels qu'entrés dans l'ordinateur), afin d'assurer la conservation des distinctions qui sont nécessaires à nos analyses. Il passe également en revue les particularités typographiques de nos corpus qui assurent, notamment, la différenciation de certains types de mots, la séparation des segments des énoncés des informateurs, et la ségrégation des énoncés des intervieweurs.

Enfin, l'auteur aborde le problème qui a motivé, en grande partie, l'informatisation : le codage des données linguistiques qui nous intéressent (en vue, bien entendu, de l'extraction automatique de telles données). Après une discussion théorique des buts, des exigences et des inconvénients de la lemmatisation (nous ne l'avons pas faite), il introduit dans sa présentation le type de codage que nous avons privilégié : nous avons codé les formes verbales en fonction du temps, du mode et de l'aspect, en employant, principalement, les chiffres de 0 à 9. L'auteur présente

également, en détail, les conventions qui déterminent et modifient la disposition des codes - conventions essentielles à la compréhension intégrale de nos corpus, tels qu'ils sont publiés ici. Aussi, nous avons exploité toutes les distinctions présentées dans cette partie de l'article (et arrêtées dans cette version des corpus) dans notre série d'études sur le système verbal des apprenants du français - série terminée et complétée par les articles de N. BAZERGUI, ici-même.

Les deux articles complémentaires de N. BAZERGUI examine certains détails importants de l'usage du système verbal du français, achevant ainsi l'examen de l'aspect de la compétence grammaticale des apprenants que nous avons privilégié dans cette recherche (voir aussi Connors et coll. 1988). En premier lieu, l'auteur aborde la grammaticalité contextuelle (syntaxique et sémantique) des choix de formes, du point de vue du temps, du mode et de l'aspect. Le codage des formes verbales dans les corpus informatisés de nos apprenants a facilité et systématisé cette partie de l'analyse (voir les articles de CONNORS et de LENOBLE dans ce recueil).

La comparaison détaillée de la performance linguistique des anglophones et des lusophones (voir la présentation des corpus ci-dessus) permet de juger si l'utilisation des temps, des modes ou des aspects reflètent l'interférence de la langue maternelle, puisque le portugais et l'anglais manifestent des chevauchements différents avec le système verbal du français. Dans les cas où ces recoupements sont liés à des facteurs sémantiques complexes et parfois obscurs (c'est le grand problème de la répartition des formes du passé), l'auteur pose l'hypothèse de la «confusion» - la substitution réciproque de deux formes (p. ex. : le passé composé pour l'imparfait et l'inverse).

L'auteur compare, tout particulièrement, l'exactitude dans l'usage du passé composé, de l'imparfait et du conditionnel, à la lumière des hypothèses d'«interférence» et de «confusion». L'infirmation partielle des hypothèses, ainsi que certaines similarités de performance entre les anglophones et les «allophones» (voir description des corpus ci-dessus) laissent croire que le système verbal pose, dans l'ensemble, des problèmes analogues à tous les apprenants du français L2.

Devant l'échec partiel des deux hypothèses de départ, l'auteur en pose une troisième : celle de l'«abdication» face à la complexité du système, menant au remplacement des formes verbales à l'étude par le présent de l'indicatif ou par l'infinitif. Les résultats de l'analyse indiquent que nos groupes d'apprenants ont tous recours à ces techniques d'«abdication», mais ils ne les emploient pas tous de la même façon et au même degré.

Finalement, l'auteur examine l'usage, standard, non-standard et non-natif, des auxiliaires *avoir* et *être* aux temps composés, en comparant nos groupes d'apprenants à la norme du français standard et à celle du français parlé à Montréal, avec des résultats intéressants qui différencient la performance linguistique des lusophones de celle des autres groupes d'apprenants.

L'auteur conclut en soulignant les limites de son analyse : elle se situe à un niveau trop général pour soulever les problèmes particuliers aux apprenants d'un groupe linguistique, ou particuliers à des contextes spécifiques de performance linguistique. C'est pourquoi elle propose, dans un deuxième temps, l'examen d'une structure en particulier, la construction verbale pronominale, celle des verbes dits «pronominaux». Non seulement cette structure est typique du français (ou, plus généralement, des langues romanes); elle est aussi requise à des degrés différents selon le thème du discours. Pour cette raison, l'auteur aborde la comparaison de la performance des anglophones et des lusophones, face aux exigences de la construction verbale pronominale, par l'intermédiaire d'une comparaison entre chaque corpus d'apprenants et le corpus francophone correspondant (voir présentation des corpus ci-dessus).

Dans un premier temps, l'auteur critique certains aspects des traitements existants des «verbes pronominaux». Les différents classements ne se recoupent qu'imparfaitement et laissent des zones grises à l'analyste qui voudrait classer de nouveaux exemples. L'auteur se propose de distinguer les cas (réfléchis ou réciproques) où le *SE* est sémantiquement motivé des cas de lexicalisation, en posant des hypothèses divergentes quant à la performance des apprenants anglophones et lusophones à l'égard de la construction verbale pronominale.

Ensuite, elle examine les occurrences de ces constructions, en comparant leur fréquence relative chez les apprenants et les francophones. Plus précisément, elle compare le corpus anglophone au corpus francophone correspondant : celui des «francophones Balablok», et le corpus lusophone au corpus «Thibault» (voir la description des corpus ci-dessus). Cette démarche permet de déterminer si la sous-exploitation apparente du système par les anglophones est tributaire des exigences du protocole d'interview (voir ci-dessus) ou non. La confrontation des corpus comparables indique, en effet, qu'il y a des différences dans l'exploitation du système des «verbes pronominaux» qui ne relèvent pas des exigences divergentes des entrevues.

Enfin, la comparaison des types d'erreurs dont la construction verbale pronominale est atteinte chez les anglophones et chez les lusophones distingue nettement leurs niveaux de maîtrise du système et permet de cerner les difficultés typiques des anglophones en particulier.

Les deux articles de B. MAJKRAK comparent l'utilisation des principaux procédés de cohésion discursive dans les deux corpus à l'étude qui se prêtent le plus à une telle démarche, les corpus informatisés provenant des entretiens portant sur le film «Balablok», ceux des jeunes apprenants anglophones et ceux des jeunes francophones du même âge et du même niveau scolaire (voir description des corpus ci-dessus).

Le premier article présente, tout d'abord, les modèles de pronominalisation déployés par les francophones et par les anglophones : les types syntaxiques de coréférence assurés par le procédé même (c'est-à-dire les fonctions grammaticales des éléments coréférentiels), les erreurs (morpho-syntaxiques) dans la réalisation des formes pronominales, et finalement les occurrences de la pronominalisation multiple (assurant la coréférence avec des SN différents).

Dans un deuxième temps, l'auteur examine l'éloignement relatif des formes pronominales de leurs antécédents, en termes des «séparateurs d'unités de base» qu'il faut franchir pour apparier les coréférents. C'est-à-dire que l'auteur différencie, *grosso modo*, le cas des SN avec pronoms coréférentiels dans une même phrase et

trois types d'anaphore à distance : dans la phrase suivante, deux phrases après, et trois phrases après ou plus.

La compilation et l'analyse des données permettent à l'auteur de démontrer l'existence de certaines différences importantes entre les anglophones et les témoins francophones, pour ce qui est de l'utilisation du système de coréférence pronominale. Nous reviendrons, à la fin de ce recueil, sur les conclusions les plus générales que nous tirons de tels écarts importants.

Dans la seconde partie de sa contribution, Majkrak se penche sur quatre autres procédés cohésifs : la paraphrase, la récurrence, le parallélisme (ou récurrence syntaxique) et l'ellipse (récupérable). Au préalable, elle a identifié les «thèmes» abordés par chaque groupe de locuteurs (anglophones et francophones). À la suite de cette étape, elle a relevé les occurrences de chacun des quatre procédés cohésifs, à l'intérieur de chaque thème, ou reliant deux thèmes, à condition qu'ils soient à la fois contigus et apparentés (sémantiquement).

La compilation et l'analyse des données révèlent des différences quantitatives importantes entre anglophones et francophones (en termes absolus, relatifs, ou les deux) dans l'apparition des thèmes pertinents au film «Balablok», et dans le taux d'occurrence de chacun des quatre procédés cohésifs. Dans la synthèse cidessous, nous reviendrons sur les implications de telles différences, qui nous semblent significatives, surtout parce que ces deux groupes de locuteurs, apprenants anglophones et témoins francophones, exécutaient la même tâche verbale pendant les entrevues qui constituent leurs corpus respectifs.

Le dernier article, celui de K. CONNORS, s'inspire des grands thèmes respectifs des deux années du projet de recherche : l'acquisition du système verbal et la cohésion du discours. Plus particulièrement, l'auteur aborde une étude des corrélations parmi les résultats d'une série d'évaluations du français des trois groupes d'apprenants à l'étude : les anglophones, les lusophones et les «allophones» (voir description des corpus ci-dessus).

L'auteur exploite trois séries d'évaluations globales de ces apprenants effectuées par des francophones, en distinguant des

évaluations «naïves», subies par les anglophones et les lusophones, des évaluations «non naïves», qui avaient appliqué la procédure d'évaluation du "Foreign Service Institute" des États-Unis. L'objectif principal de l'article est la comparaison des corrélations obtenues entre deux types d'indices linguistiques, d'une part, et l'évaluation globale, «naïve» ou «non naïve», d'autre part. Si l'on suppose que l'évaluation globale de la performance de l'apprenant par un locuteur natif puisse servir de vérification externe à une évaluation purement linguistique, on pourra considérer le degré de correspondance entre les évaluations globales et les différentes évaluations linguistiques comme test des critères linguistiques sous-jacents, afin de déterminer lequel de ces derniers mérite le statut d'indice du niveau d'acquisition de la L2.

Dans ce sens, l'étude met à l'épreuve, essentiellement, deux mesures linguistiques, l'une basée sur la grammaticalité (en contexte) du choix, par l'apprenant, des formes verbales, en termes de temps, de mode et d'aspect, et l'autre basée sur la variété lexicale (relative) des connecteurs et des incises. (Ces dernières sont des expressions courtes et figées qui s'insèrent à l'intérieur ou à la fin des propositions, dans le contexte d'un discours). La discussion tourne autour de la vérification des hypothèses de l'auteur, qui a prédit une corrélation étroite entre l'évaluation «naïve», d'un côté, et la mesure de variété des connecteurs et des incises, de l'autre; l'évaluation «non naïve» correspondrait plus étroitement aux résultats de la mesure de la grammaticalité des formes verbales en contexte. Malgré ses justifications *a priori* de ces hypothèses, l'auteur constate que le comportement linguistique des lusophones les infirme, même si la performance de deux autres groupes les confirme. La conclusion de l'article traite essentiellement de cette contradiction apparente, et de ses implications pour l'établissement de corpus d'apprenants qui sont basés sur des entrevues «semi-dirigées».

Références

Andersen, Roger (éd.). (1984). *Second languages. A cross-linguistic perspective*. Rowley, MA : Newbury.

Berthoud, A. (1982). *Activité métalinguistique et acquisition d'une langue seconde*. Berne : Lang.

Brown, Roger. (1973). *A first language*. Cambridge, MA : Harvard University Press.

Clahsen, H., J.M. Meisel et M. Pienemann. (1983). *Deutsch als Zweitsprache. Der Spracherwerb ausländischer Arbeiter*. Tübingen : Narr.

Connors, K., M-A. Vieira, J. Reighard et M. Dillinger. (1988). "Le français des lusophones à Montréal : particularités morphologiques des variétés discursives", in Kremer, D. (éd.). *Actes du XVIII^e Congrès international de linguistique et de philologie romanes*, tome V, Tübingen : Niemeyer, pp. 307-19.

D'Anglejan, A. et coll. (1986). "Beyond the language classroom : a study of communicative abilities in adult immigrants following intensive instruction". *TESOL Quarterly* 20 : 2, 185-205.

D'Anglejan, A. et C. Renaud. (1985). "Learner characteristics and second language acquisition : A multivariate study of adult immigrants and some thoughts on methodology". *Language Learning* 35 : 1, 1-19.

Dechert, H.W., D. Möhle, et M. Raupach (éds). (1984). *Second language productions*. Tübingen : Narr.

Deulofeu, J. et C. Noyau. (1986). "L'étude de l'acquisition spontanée d'une langue étrangère : méthodes de recherche/méthodes en linguistique/apports". *Langue Française* 71, 3-16.

Elbaz, Charles. (1980). "Les Centres d'Orientation et de Formation des Immigrants : C.O.F.I. Origines - Évolution - Situation actuelle". Faculté des Sciences de l'Éducation, Université de Montréal.

Ellis, Rod. (1986). *Understanding second language acquisition*. Oxford : Oxford University Press.

Genesee, Fred. (1987). *Learning through two languages : studies of immersion and bilingual education*. New-York : Newbury.

Giacobbe, J. et M.-A. Cammarota. (1986). "Learners' hypotheses for the acquisition of lexis". *Studies in Second Language Acquisition* 8 : 3, 327-42.

Giacobbe, J. et M. Lucas. (1982). "Métalangue des apprenants et discours spontané : connaissances explicite et implicite dans l'acquisition des verbes espagnols *ser* et *estar*", in Trévise, A. (éd.). *Encrages*. Paris : Université de Paris VIII, pp. 111-27.

Giacomi, A. et D. Véronique (éds). (1986). *Acquisition d'une langue étrangère. Perspectives et recherches*. Tomes I et II. Actes du 5^e Colloque international, Aix-en-Provence 1984. Université de Provence.

Grandcolas, B. et N. Soulé-Susbielles. (1986). "The analysis of the foreign language classroom". *Studies in Second Language Acquisition* 8 : 3, 293-308.

Holobow, N.E. et coll. (1987). "Effectiveness of partial French immersion for children from different social class and ethnic backgrounds". *Applied Psycholinguistics* 8 : 2, 137-52.

Klein, W. et C. Perdue (éds). (1988). *Utterance structure*. Strasbourg et Nimègue : European Science Foundation.

Langue Française 71 (1986). Paris : Larousse.

Lightbown, P.M., N. Spada et coll. (1988). "Intensive ESL in Quebec". Communication présentée au "TESL Canada Summer Forum".

Möhle, Dorothea. (1984). "A comparison of the second language speech production of different native speakers", in Dechert, H.W. et coll. (éds), pp. 26-137.

Perdue, C. (1987). "Understanding and misunderstanding in adult language acquisition : Recent work on the E.S.F. Project", in Lüdi, G. (éd.). *Devenir bilingue, parler bilingue*. Tübingen : Niemeyer.

Perdue, Clive (éd.). (1984). *Second language acquisition by adult immigrants. A field manual*. The European Science Foundation project. Rowley, MA : Newbury.

Perdue, C. et R. Porquier (éds). (1979). *Encrages : Numéro spécial de linguistique appliquée*. Université de Paris VIII.

Perdue, C. et R. Porquier (éds). (1980). "Apprentissage et connaissance d'une langue étrangère". *Langages* 57. Paris : Larousse.

Perdue, C. et coll. (1986). "L'acquisition du français par des adultes immigrés". *Langages* 84. Paris : Larousse.

Perdue, C., J. Deulofeu et A. Trévise. (1988). "The acquisition of French", in Klein, W. et C. Perdue (éds), pp. 234-313.

Pienemann, Manfred. (1987). "Is language teachable? Psycholinguistic experiments and hypotheses". *Arbeiten zur Mehrsprachigkeit* 21, 1-32.

Piétropaolo-Saura, B. et M. Roffé, (1985). "Analyse différentielle et analyse des erreurs : un nouveau regard". *Bulletin de l'ACLA* 7 : 1, 47-55.

Py, Bernard (éd.). (1984). *Acquisition d'une langue étrangère III*. Actes du colloque organisé les 6-18 septembre 1982 à l'Université de Neuchâtel. Université de Neuchâtel.

Raupach, Manfred. (1984). "Formulae in second language speech production", in Dechert, H.W. et coll. (éds), pp. 114-37.

Singleton, David. (1987). "Mother and other tongue influence on learner French : a case study". *Studies in Second Language Acquisition* 9 : 3, 327-45.

Singleton, David. (1986). *Studies in second language acquisition* 8 : 3. New-York : Cambridge University Press.

Towell, Richard. (1987). "Variability and progress in the language development of advanced learners of a foreign language", in Ellis, Rod (éd.). *Second Language Acquisition in Context*. Englewood Cliffs N.J. : Prentice-Hall International.

Trévise, A. (éd.). (1982). *Acquisition d'une langue étrangère II* (*Encrages* 8/9). Université de Paris VIII.

Trévise, A. et C. Noyau. (1984). "Adult Spanish speakers and the acquisition of French negation forms : individual variation and linguistic awareness", in Andersen, R. (éd.), pp. 165-89.

Trévise, A. et R. Porquier. (1986). "Second language acquisition by adult immigrants : exemplified methodology". *Studies in Second Language Acquisition* 8 : 3, 265-75.

Véronique, Daniel. (1984). "The acquisition and use of aspects of French morphosyntax by native speakers of Arabic dialects (North Africa)", in Andersen, R (éd.), pp. 191-213.

White, Lydia. (1987). "Markedness and second language acquisition : the question of transfer". *Studies in Second Language Acquisition* 9 : 3, 261-86.

Transcription, encodage et normalisation de corpus oraux en vue d'une exploitation linguistique semi-automatisée

Michel Lenoble

Introduction

L'ensemble de nos travaux de recherche s'est déroulé sur une période de trois années, dont la première a été largement consacrée à l'établissement des hypothèses de recherche et des indices linguistiques permettant le mieux de les illustrer. Parallèlement à cette activité, d'autres énergies étaient consacrées à la collecte de donnés empiriques sur le terrain. Nous publions ces trois corpus informatisés dans le volume II : *Acquisition du français (L2) chez des adultes à Montréal : corpus informatisés : anglophones, francophones et lusophones*. Deux de ces corpus résultent d'entrevues faites avec des locuteurs du français L2. Ces deux groupes cibles d'allophones représentent deux types de clientèles particulières et traditionnelles d'apprenants du français L2 au Québec. Le premier groupe est constitué de locuteurs lusophones européens, immigrés à Montréal. Le deuxième groupe est composé de jeunes locuteurs anglophones du Québec. Le troisième corpus présente des entrevues avec des jeunes locuteurs francophones «témoins», qui ont subi la même entrevue que celle des jeunes anglophones. Voir la description, plus détaillée, de l'origine de l'ensemble des corpus à l'étude dans l'*Introduction* ci-dessus.

Nous voulons exploiter nos trois corpus informatisés de façon à la fois séparée et particulière et à la fois de façon contrastive, en nous servant de chacun des trois corpus successivement comme objet d'étude et comme repoussoir afin de mettre nos résultats en perspective.

Étant à l'affût d'indices très ténus et dissimulés sur l'ensemble des 54 entrevues nous permettant de mettre au jour des particularités propres à chacun des niveaux d'apprentissage du français L2 pour les deux groupes cibles et de mener à bien des études fines sur la correction de l'usage de la richesse du système verbal français, ainsi que sur la cohérence et la cohésion du lexique, il nous est apparu intéressant, au début de la deuxième année du projet, de procéder à l'informatisation de nos données. De toute évidence, c'est toujours une tâche lourde et fastidieuse si nous voulons nous assurer d'un travail de qualité, mais néanmoins combien indispensable lorsqu'il s'agit d'avoir sur une énorme masse de données[1] une emprise totale et un accès exhaustif aussi rapide que dépourvu d'erreurs d'inattention. Nous traiterons dans les pages qui vont suivre des diverses étapes du processus d'informatisation menant des versions enregistrées des corpus aux textes sur support informatique. Nous expliquerons les choix que nous avons dû faire en raison des limites inhérentes au progiciel d'exploitation et les normes que nous nous sommes fixées pour l'établissement de l'ensemble des corpus et enfin les indices et la manière de les repérer que nous avons mis au point, afin de faire ressortir les traits linguistiques pertinents pour nos recherches.

Les corpus enregistrés

Les entrevues avec les sujets lusophones se sont déroulées sur un mode semi-directif; en effet, la conversation ne partait pas d'événements précis ou ne survenait pas à la suite d'une mise en situation préarrangée. Les réponses sont donc de longueurs variées et les questions respectaient un enchaînement naturel, puisque l'interrogateur réagissait spontanément à ce qui venait d'être dit et ne proposait pas un choix limité de réponses possibles. Les thèmes de conversation tournaient autour du récit de souvenirs d'enfance de ces locuteurs, de la relation de leur arrivée au Québec et des débuts de leur installation, ainsi que de leur adaptation à la vie nord-américaine, leur récit de vie se terminant par la description

d'un jeu, d'un conte, ou d'une recette de cuisine et de ce qu'ils feraient en cas de gain à la loterie.

À l'opposé, les corpus francophone et anglophone ont été réunis suite à une mise en situation, puisqu'il s'agissait, pour les locuteurs, après le visionnement d'un film animé, de raconter ce qu'ils en avaient retenu et de donner une esquisse de l'interprétation qu'ils en faisaient. Ici donc, il y avait un protocole de questionnement plus strict, mais il ne proposait pas de réponses possibles, exigeant, au contraire, une certaine autonomie du répondant, qui devait justifier ses réponses.

Cependant, les trois collectes de données se sont déroulées de façon à ce que les sujets soient amenés à produire un certain nombre d'énoncés dont les constituantes «grammaticales» nous intéressaient. Premièrement, le canevas général de chaque entrevue devait inciter le sujet à produire trois types de discours, à savoir le discours narratif relatant des faits passés, le discours descriptif reflétant des actions présentes et le discours hypothétique leur permettant d'élaborer des prédictions concernant des éventuels événements futurs. Chaque type de discours met en jeu, actualise des potentialités du langage qui lui sont typiques dont l'une des plus révélatrices serait l'usage des temps, des aspects et des modes. Les questions ainsi posées devaient susciter, de la part des sujets, l'emploi de temps précis et les écarts par rapport à l'usage courant devaient nous permettre d'évaluer le niveau d'apprentissage ou bien la qualité de la maîtrise du français L2.

En ce qui concerne l'étude de la cohésion lexicale, elle est aussi rendue plus attrayante par le fait que les corpus se subdivisent en entrevues séparées par sujet, ce qui permet à la fois des études de cas particuliers et des comparaisons par rapport à la moyenne du groupe. Les deux corpus constitués à la suite du visionnement du même film animé se prêtent en plus à une comparaison translinguistique, puisque francophones et anglophones ont été appelés à réagir à la même mise en situation.

Transcription manuscrite

Le passage aux données sur support informatique se fait généralement à partir de données écrites, fussent-elles manuscrites

ou imprimées; il y a néanmoins un certain nombre d'étapes préparatoires à suivre avant de procéder à l'entrée des données. Dans le cas de corpus oraux, sur support magnétique, la transition par l'étape de l'écrit, sans être obligatoire, facilite grandement la prise de décision concernant deux aspects distincts : la première facette du travail est d'évaluer la version orale : il s'agit de distinguer ce que l'on entend et de décider comment on le transcrit. La deuxième tranche de ce travail préparatoire sera de sélectionner les codages nécessaires pour normaliser la mise en page des corpus et pour adapter les données aux impératifs du progiciel d'exploitation que nous avons choisi. Toutes ces raisons font qu'il est difficile d'escamoter la phase de l'écrit.

Le défi est de transcrire de façon optimale tout ce que l'on entend sur les rubans magnétiques, ni plus ni moins. Tout d'abord, certains passages des entrevues sont inaudibles et chaque fois que tel était le cas nous avons représenté les passages manquants par des points de suspension. Le problème s'est posé presque uniquement pour les questions posées par les intervenants, étant donné que le micro utilisé lors des entrevues était de type unidirectionnel et non un micro d'ambiance ou cardioïque, qui nous aurait permis de percevoir également les paroles des personnes se trouvant en arrière du micro. D'autres portions de texte présentaient une prononciation très éloignée de la prononciation standard, étaient constituées de mots n'appartenant ni au français, ni à la langue maternelle des sujets. Ces bribes étaient étymologiquement proches de leur langue maternelle et phonologiquement proche du français (p. ex. : /rejection/). Enfin, nous avons aussi rencontré, en proportions très variables d'un groupe à l'autre et d'un sujet à l'autre, des mots de leur langue maternelle en remplacement de mots du français qu'ils étaient incapables de prononcer (p. ex. : /hectagone/), qu'ils ne connaissaient pas ou dont ils ne pouvaient se souvenir.

Ces trois catégories de mots, à prononciation non-standard, à morphologie et prononciation hybride ou appartenant à leur langue maternelle, ont été marqués dans la transcription par l'adjonction d'une barre oblique avant et après. Ceci nous permet à la fois de les distinguer facilement à la lecture et de les regrouper aisément à la fin des extrants.

L'hypercorrection est en soi un phénomène intéressant linguistiquement; par contre, il est beaucoup moins intéressant de constater que certains des transcripteurs de la première heure ont plus d'une fois été tentés de corriger inconsciemment la réponse des sujets lors de leur travail de transcription, ce qui avait pour effet dévastateur d'oblitérer complètement ce que nous étions très précisément en train d'essayer de repérer. Nous avons donc dû restituer les «fautes» à leurs auteurs suite à une seconde écoute-correction des enregistrements.

Restait également le cas des homophones, auxquels il s'agissait de donner la bonne graphie; nous l'avons fait en tenant compte de deux principes : là où plusieurs graphies existent pour rendre une chaîne phonique, nous avons opté pour la désambiguïsation par contexte minimal, cette technique nous permettant de choisir l'orthographe appropriée en fonction du contexte morphologique ou syntaxique, ou bien en regardant les mots auxquels la chaîne en question était associée par coordination. Par exemple, pour arrêter la graphie d'un verbe de la première conjugaison se terminant par un son /é/ il suffisait de voir s'il était coordonné à un autre verbe d'une autre conjugaison à l'infinitif ou s'il était précédé d'un auxiliaire. À chaque fois que le contexte minimal ne nous permettait pas de trancher, nous avons donné régulièrement le bénéfice du doute au locuteur et nous lui avons accordé la transcription qui était la plus correcte. Par exemple, dans la phrase «Il faut qu'il *mette*7 de l'ordre dans ses affaires», nous avons indiqué la forme verbale au subjonctif et elle a été codée comme telle, plutôt que de l'indiquer comme la troisième personne du présent de l'indicatif entre barres obliques /met/.

La deuxième tranche de cette phase préparatoire consiste en la normalisation des corpus et en l'adaptation des codages au progiciel. L'absence de version écrite ou imprimée de référence nous donne un peu plus de liberté pour exécuter cette partie du travail.

En règle générale, la version magnétique du corpus correspond toujours, quant à sa présentation physique, à la version typographiée, en présentant les mêmes fins de ligne, les mêmes marques de paragraphe et de fins de page, afin de permettre un retour facile et immédiat à la version papier. Cette dernière n'existant ici que

lors de la transcription manuscrite intermédiaire, qui n'est jamais rendue publique, nous avons tout le loisir de fixer à notre gré la mise en page que la copie papier de notre corpus sur support informatique prendra.

Dans le cas qui nous occupe, nous avons donc créé de toutes pièces la version de référence en nous efforçant d'en fixer l'aspect et la forme, en respectant les paramètres qui devaient nous servir lors de la fabrication des prêt-à-photographier et de la publication sous forme imprimée.

Pour des raisons tant matérielles qu'informatiques, la longueur des lignes a été fixée à 58 caractères et le nombre de lignes par page est de 57. Toutes les entrevues présentent ainsi la même mise en page. La disposition typographique présente encore une particularité, puisque nous avons la possibilité, avec le progiciel utilisé, d'introduire au début de chaque ligne du corpus un certain nombre d'indications devant servir à faciliter la référenciation ou le retour à la copie papier. En effet, le progiciel nous fournit, entre autres, des listes de vocabulaire triées selon des ordres différents et indique pour chacune des occurrences l'adresse où l'on peut retrouver le mot dans le corpus. Ainsi, le champ libre est constitué de 52 caractères de texte et le champ fixe, situé à sa gauche, indique, sur une longueur de 6 caractères, le numéro de la page et de la ligne[2]. Ce sont justement ces chiffres qui sont repris comme adresse pour les mots se trouvant sur la ligne.

Pour faciliter la lecture des corpus, nous avons séparé nettement les questions des interviewers du reste du corpus en les mettant en évidence en disposant une ligne blanche avant et après chaque question, dont le libellé est d'ailleurs aussi souligné[3]. Les questions sont aussi les seules phrases du corpus qui commencent par une lettre majuscule. Cette distinction typographique se trouve aussi répercutée au niveau du progiciel, puisque nous avons la possibilité de distinguer, dans le texte des données, plusieurs «langues». Le concept de langue est pris dans un sens bien particulier par notre logiciel, puisqu'il recouvre tout rassemblement de codes informatiques précédés par un identificateur ou embrayeur de langue précis et suivis par l'embrayeur d'une autre langue. De cette façon, nous sommes par exemple à même de regrouper tous les mots prononcés par chacun des personnages d'une pièce en

des rassemblements lexicaux distincts. Le même mécanisme permet de la même façon d'isoler, si on le désire, les mots mis en évidence par un choix de fonte typographique particulière en les marquant par des embrayeurs. Pour nos entrevues, les mots du lexique des interrogateurs n'avaient en soi qu'une fonction de stimulation et d'orientation de la conversation et ne devaient en rien influencer le décompte de vocabulaire entrepris pour les groupes cible. C'est pourquoi nous avons eu recours au procédé d'isolation de parties de corpus de données par l'adjonction d'embrayeurs de langue pour séparer les mots des questions des mots des réponses. Nous nous sommes servi de deux codes non utilisés, soit $ et %, pour marquer le début et la fin de la langue question. La même technique et la même mise en page ont été utilisées pour les trois corpus.

Entrée/validation des données

L'entrée des données se déroule en plusieurs étapes, dont la première consiste à rentrer presque exclusivement le texte et un minimum de codes. Chaque entrevue est contenue dans un fichier informatique distinct et si le besoin de traiter l'entièreté d'un corpus simultanément se fait sentir, le système d'exploitation du système informatique nous laisse le choix entre plusieurs options élégantes pour relier les fichiers entre eux, voire même de les fusionner. Ainsi, toute recherche ponctuelle peut être menée soit sur l'entrevue d'un seul sujet à la fois, soit sur l'ensemble du corpus.

Pendant la deuxième phase de l'entrée des données nous nous cantonnons à n'ajouter que les codages de mise en page et de séparation des différentes «langues» (question et réponse) du texte.

Ces phases d'entrée alternent avec des phases de relecture et de correction visant toujours un seul aspect particulier à la fois, soit les codages pré-cités, soit les embrayeurs, etc. Dans ces tâches fastidieuses nous nous servons de la puissance du logiciel d'édition de texte supporté par les CYBER du Centre de Calcul de l'Université de Montréal, afin de mettre en évidence l'élément particulier qui est l'objet de notre attention lors de la correction[4]. Nous utilisons aussi le logiciel JEUDEMO (voir ci-dessous) pour dépister les codages ou les oublis dans les premières versions des fichiers

informatiques ainsi constitués. La correction d'aspects particuliers du codage passe la plupart du temps par le dépouillement scrupuleux de nombreux extrants.

Il s'agit ensuite de valider l'ensemble des caractères non orthographiques comme les signes de ponctuation et de vérifier la présence des codes nécessaires pour représenter les symboles diacritiques (p. ex. : le tréma est représenté par la voyelle suivie du signe =; p. ex. : *égoi=ste*, voir a.17.4.20). Avant de procéder à des corrections lors de lectures cursives des corpus, il reste à étudier le cas des mots composés et que nous désirons garder tels lors de notre exploitation des corpus; il s'agit, entre autres, des mots comportant un trait d'union, une apostrophe de liaison ou des expressions idiomatiques constituées de plusieurs mots où les «blancs» sont considérés comme des espaces de liaison concatenant ces divers mots. Ces derniers doivent être traités comme des unités ne formant qu'un bloc.

En fin de processus d'entrée, il reste à ajouter les codages des mots phonétisés ou étrangers, ainsi que tous les codages indispensables pour faire ressortir les éléments du texte que nous recherchons, à savoir ici, dans le cas des deux corpus d'apprenants, les codes de temps (mode, aspect) et les codes de formes verbales présentant une erreur de choix de temps (mode, aspect).

Pour terminer le processus de validation des données, il faut également éliminer les fautes de frappe inévitables. Les fautes de frappe sont inhérentes à tout travail de dactylographie, mais deux paramètres principaux font varier très fortement leur fréquence. La qualité du matériel informatique des périphériques d'entrée : dans notre cas, la malchance a voulu que nous entrions nos corpus sur des terminaux vétustes, qui devaient être - nous l'avons appris à nos dépens - remplacés quelques temps plus tard. La qualité du travail s'est pour cette raison avérée très faible; en effet, d'innombrables caractères étaient dupliqués involontairement alors que d'autres n'apparaissaient pas du tout. Deuxièmement, en plus de l'apprentissage progressif de l'utilisation efficace de l'éditeur de texte, la multiplicité des tâches à accomplir simultanément (p. ex. : entrée de texte, de codage de mise en page, de codage des verbes, des diacritiques, etc.) est un des éléments faisant varier de beaucoup le taux d'erreur de frappe. Nous avons pu réduire ce

facteur au minimum en sectionnant la phase d'entrée en plusieurs tronçons autonomes.

Sans vouloir pondérer l'influence directe de chacun de ces facteurs, nous pouvons néanmoins conclure que nous avons eu un taux d'erreurs plus élevé que la normale à cause de la complexité de la tâche et surtout à cause du facteur «matériel». Le temps supplémentaire passé à apporter tout près de 10 000 corrections en tous genres nous a toutefois permis de connaître notre corpus de données, ses reliefs et ses contours en profondeur.

Normalisation des fichiers

La normalisation consiste en la vérification que les travaux d'entrée et de codage aient bien respecté, pendant toute la durée des opérations, toutes les normes que l'on s'était fixées au départ. Ainsi nous avons vérifié que les majuscules soient uniquement utilisées dans la langue «question» au début des phrases. Les seuls mots qui comptent une majuscule dans la langue réponse sont les noms propres et quelques acronymes. La décision d'agir ainsi a été prise vu le fait que nous ne voulions nous engager à marquer la fin des phrases par une quelconque ponctuation. Les seules marques qui indiquent un tant soit peu la cadence ou le débit de la parole et ses interruptions sont les tirets, figurant une pause longue ou une hésitation marquée de la part du locuteur. Nous avons donc omis les majuscules pour ne pas imposer au texte une division par trop rigide ni ouvrir le débat des énoncés incomplets et des propositions incises, bien qu'il y ait là matière à plus sérieuse prospection.

L'alternance des questions et des réponses à été clairement balisée par des embrayeurs de langue, la première lettre majuscule, le soulignement du libellé et le saut de ligne entre les portions de texte des locuteurs différents. Même les mots isolés, la plupart du temps de simples suggestions de mots de vocabulaire posant problème au sujet, ont été mis en évidence de la même façon que les questions complètes. Donc aucun des mots de la langue «question» n'est mélangé avec les mots de la langue «réponse» lors des recherches ou des décomptes de vocabulaire. En plus, les majuscules sont pour le logiciel JEUDEMO de deux types qu'il distingue aisément lors du traitement; les majuscules de début de

phrase sont écrasées et donc réduites à l'état de minuscules quand il s'agit de comptabiliser le nombre de formes différentes dans un texte ou la fréquence d'un mot, que sa forme soit majuscule ou minuscule.

La mise en page est aussi fixée une fois pour toutes, puisque cette dernière ne peut varier continuellement, étant donné le système de référenciation que nous utilisons avec le champ fixe servant d'adresse pour repérer dans le texte les mots repris dans les extrants. C'est aussi cette adresse dont les membres de l'équipe se servent quand ils font des renvois au matériel dans le cours de leurs articles. C'est pourquoi la forme doit être standardisée et définitive.

En termes informatiques, les lettres accentuées ne posent plus de problème, puisqu'elles sont de nos jours facilement affichées à l'écran ou générées sur imprimante et que, de plus, elles sont déclarées, lors des tris alphabétiques, équivalentes à la lettre correspondante non accentuée; ainsi, les mots sont triés selon l'ordre de déclaration des symboles orthographiques à l'intérieur d'une même équivalence.

Le signe typographique «-» possède en français plusieurs sens distincts selon les situations où il est utilisé. Tous ses usages, sauf un, ne requièrent aucun codage supplémentaire dans l'enregistrement de nos corpus. Seul le tiret de liaison, reliant deux composantes d'un mot composé, est représenté de façon particulière, puisqu'il est redoublé afin d'indiquer au logiciel que les groupes de caractères le précédant et le suivant doivent être considérés comme une seule unité et traités comme tel lors de l'exécution des programmes (p. ex. : peut--être).

Le redoublement de l'apostrophe procède de la même logique et indique également que la chaîne de caractères doit être conservée d'un seul tenant (p. ex. : aujourd''hui). Bien que nous soyons totalement libres d'utiliser la possibilité de liaison où et quand nous le désirons, que ce soit pour le tiret, l'apostrophe ou l'espace, nous avons suivi sur toute l'étendue des corpus la même logique de codage, afin de rassembler uniformément les unités lexicales.

Le progiciel d'exploitation

JEUDEMO (voir ci-dessus) est un progiciel d'exploitation de données textuelles, appartenant aux langages de programmation de quatrième génération, qui a été rédigé au début des années 80 par Francine Ouellette[5] et supporté depuis par le Centre de Calcul de l'Université de Montréal. Il permet de produire des listes de mots, des concordances et des statistiques élémentaires sur les textes. Le logiciel accepte des textes d'alphabet, de mise en page et de taille variés.

JEUDEMO est un langage de programmation dont les mots-clé appartiennent à notre langue naturelle; sa syntaxe est stricte et vérifiée lors de la soumission par le compilateur du système. La structure de tout programme en JEUDEMO se divise en trois parties[6]. Il nous appartient d'abord de décrire le texte à étudier; ensuite, nous devons préciser la ou les commandes désirées et enfin il reste à spécifier la disposition et le contenu des listes commandées. Ainsi l'utilisateur détermine l'ordre alphabétique approprié à la langue de son texte, l'ordre de tri des listes recherchées et l'ampleur de la portion de texte analysée. Les mots recherchés peuvent être précisés dans un intervalle alphabétique, un intervalle de fréquences ou un intervalle de longueurs, ou encore selon des sous-ensembles de mots définis par une liste de mots explicites, par des préfixes ou des suffixes. Les statistiques minimales que JEUDEMO est en mesure de produire sont : le nombre de caractères, le nombre de lettres, le nombre de mots de formes différentes, le nombre de mots total, le nombre de lignes ainsi que la fréquence absolue et la fréquence moyenne des mots.

Lors de la déclaration des symboles alphabétiques au début de la rédaction d'un programme en JEUDEMO, des symboles autres que les lettres usuelles de l'alphabet peuvent être inclus. Ceci a pour effet de modifier les résultats du programme. Dans l'étude présente, nous avons inclus les symboles de liaison afin de conserver d'un seul tenant les unités lexicales qui en contiennent. En outre, les barres obliques ont été insérées à la fin de la liste des symboles alphabétiques, ce qui implique que se retrouvent regroupés à la fin des extrants tous les mots qui débutent par une barre oblique, c'est-à-dire tous les mots dont l'origine ou la prononciation ne nous permet pas de les identifier à la langue

française. Si la barre oblique n'avait pas été incluse dans la liste alphabétique, ces mêmes mots auraient été classés parmi les autres, sans possibilité de les distinguer facilement[7]. Enfin, les embrayeurs de langue ($ et %) marquent, chacun à leur tour, le début d'une série de mots devant être comptabilisés et rassemblés sous une même bannière et représentant le vocabulaire respectif de l'interviewer ou du répondant. Rappelons également que les trémas sont représentés dans la version informatique des fichiers par la voyelle suivie du signe =.

Les codages propres à la recherche

Afin de préciser plus encore le type de chaîne de caractères qui nous intéresse particulièrement, nous pouvons ajouter des codages supplémentaires qui sont sélectionnés et positionnés dans le fichier informatique uniquement pour resserrer les critères de sélection. Il s'agit donc de définir le genre d'unités de base à partir desquelles on souhaite travailler et d'adapter les codages afin de les faire ressortir aisément.

Pour fixer clairement la terminologie de l'exposé qui suit, cinq termes désignant cinq types d'unités résultant de logiques différentes de découpage du texte doivent être définis brièvement. La partition la plus simple d'un enregistrement magnétique d'informations se réalise en se servant des signes de ponctuation ou des espaces typographiques pour isoler les formes lexicales. Fréquemment le seul critère graphique de découpage n'est pas suffisant car unités sémantiques et unités graphiques sont loin de toujours être parfaitement équivalentes. En effet, soit une unité graphique contient plusieurs unités sémantiques comme dans les articles contractés et les groupes graphiques reliés par une apostrophe ou un trait d'union (p. ex. : *de, du, l'agneau, grand-peur*); soit aussi que l'unité sémantique se compose de plusieurs unités graphiques, comme les phénomènes lexicaux tels que *tour à tour* et les phénomènes morpho-syntaxiques tels que *ayant été écoutés*. Dans un cas comme dans l'autre, le chercheur doit ajouter à son texte magnétique des codages supplémentaires afin que cette nouvelle version reflète aussi précisément que possible la logique qui relie les divers éléments constitutifs du manuscrit. Cette étape de raffinement qui nous mène de la forme au vocable corrige les anomalies résultant de la stricte application de la règle de la

séparation des formes. Donc, en distinguant dans notre programme JEUDEMO les apostrophes de liaison et de séparation, nous éviterons d'avoir deux formes différentes pour *c'était* et *était* et proposerons deux entrées pour *l'abord* et *d'abord*.

Cette dernière démarche témoignait seulement d'une exigence de rigueur et de précision basée sur la signification du mot en séparant ce qui est de même forme sans avoir le même sens. Mais si le chercheur veut satisfaire à l'exigence de compression des données afin de mieux les dominer, il se doit de lemmatiser les vocables. La logique de cette opération est opposée à celle de la précédente : elle rassemble cette fois ce qui s'écrit différemment, mais qui est de lemme commun. Cette opération se déroule sur deux fronts : 1) elle vise à réunir les vocables hétérographes sous un même lemme, comme les vocables polymorphes *(beau, bel)*, les vocables élidés *(l', le)* et toutes les formes fléchies des substantifs, adjectifs, verbes et pronoms; 2) elle vise à séparer les vocables homographes de sens différents comme dans le cas de polysémie, d'homonymie, de dérivation impropre et de changement de catégorie grammaticale. La grande diversité des cas à traiter et l'infinie variété des projets de recherche obligent à chaque fois le chercheur à se définir une norme lexicologique afin d'avoir une certaine constance dans ses choix. Retenons qu'une liste de vocables lemmatisés gagne toujours en précision sinon en condensation par rapport à une liste non-lemmatisée et présente donc une vue plus synthétique des faits lexicaux.

L'aspect sémantique est prépondérant dans la motivation du découpage des unités d'information que je viens de décrire. Des recherches plus récentes tentent aussi de fragmenter l'atome sémantique des vocables en sous-parties comptant un noyau ou radical sémantique et des ions baladeurs tels les préfixes et les suffixes. Cette division s'opère selon les règles de l'étymologie et vise à pouvoir regrouper dans des listes les vocables en fonction de radical, préfixes et suffixes. Cette opération baptisée sématisation permet de faire d'intéressants regroupements et de percevoir la logique profonde de la langue d'un texte. Cette approche qui mettrait en relation des mots comme *imbécile* et *bâton* grâce à leur sème commun *baculum* est bien sûr microscopique et infradiscursive, mais répond à un impératif d'explicitation. L'unité de base en est le sème.

Un travail supplémentaire de codage du texte, la catégorisation, permet d'accoler à chaque vocable des renseignements décrivant les catégories auxquelles il appartient. Les types de catégories les plus souvent retenues sont les classes morphologiques, syntaxiques et stylistiques. Une fois le texte entièrement catégorisé, de nombreuses listes peuvent être produites dont l'unité de base d'information sera la catégorie (p. ex. : la liste de tous les verbes d'un texte sélectionnés et triés d'après le temps, la personne, le nombre ou le mode).

La question de savoir s'il faut absolument lemmatiser un corpus ou pas avant de commencer son exploitation ne peut se poser en termes absolus. En effet, il ne s'agit pas, lors d'une recherche précise, de produire un corpus présentant des codages pouvant répondre à tous les types de recherches potentielles, mais bien au contraire d'arriver à disposer d'un texte dont le codage est suffisant pour réaliser la production automatique de tous les extrants nécessaires à la réalisation d'une recherche particulière. En soi, un travail de lemmatisation est très exigeant en temps et en ressources humaines et nécessite une réflexion théorique poussée afin d'établir une norme lexicologique cohérente. De plus, pour la recherche qui nous occupe ici, la lemmatisation est loin d'être indispensable. En effet, nous ne nous intéressons qu'à un nombre limité de traits linguistiques pertinents en tant qu'indicateurs du niveau d'apprentissage du français L2, à savoir les formes verbales. Coder uniquement les verbes présentait le double avantage, d'une part, d'en automatiser le listage et, d'autre part, de les isoler de toutes les autres formes lexicales homographes (p. ex. : *deux marches* et *tu marches*1). Ce travail minimal est en fait une catégorisation partielle du corpus puisqu'elle en présente les mêmes avantages et est conduite en respectant la même logique sous-jacente, tout en économisant beaucoup de temps.

Par ailleurs, en ne lemmatisant pas le texte, les formes verbales différentes que peut prendre un verbe (p. ex. : plus de 80 pour le verbe *être*) ne sont pas toutes automatiquement comprimées sous une seule et même entrée, leur lemme, mais bien au contraire, nous conservons intact dans nos extrants l'éventail complet des diverses formes verbales que les répondants utilisent et dont nous voulons étudier la richesse.

Les formes verbales ont donc été codées de façon systématique sur l'entièreté des corpus lusophone et anglophone. Les codes retenus sont les chiffres arabes de 0 à 9 et le signe *#*. Partout dans le déroulement des entrevues où des nombres sont mentionnés, ils sont transcrits en toutes lettres pour éviter la confusion éventuelle avec les codages des verbes. Les chiffres arabes employés sont placés à la fin du verbe et y sont accolés. Le chiffre 1 correspond au présent, le 2 à l'imparfait, le 3 au passé composé, le 4 au plus-que-parfait, le 5 au participe, le 6 au futur, le 7 au subjonctif, le 8 au conditionnel, le 9 à l'infinitif, le 0 au futur périphrastique et le *#* à l'impératif. Afin de nous donner plus de souplesse dans le codage de nos verbes, nous avons redoublé le code du temps dans le cas où ce dernier était à une forme passée, comme par exemple 88 est attribué au conditionnel passé. En ce qui concerne les formes des temps composés, les auxiliaires n'ont pas été codés, seul le verbe l'est. Le même principe a été suivi pour les formes passives dont seul le verbe s'est vu accoler le code prévu pour la forme active correspondante. Dans le cas des locutions verbales composées d'un verbe précédé d'un modal, les deux composantes ont été codées, les deux ayant été considérées comme verbes pleins. Les locutions verbales composées d'une forme du verbe *aller* et d'un infinitif comme *il va voyager cet été* ont été codées futur périphrastique. Par contre le même type d'expression peut aussi dénoter une habitude comme dans la phrase *Le matin il va déjeuner au coin*; dans ces cas, nous avons codé les verbes *va1* et *déjeuner9*. L'intérêt de ne pas avoir codé les auxiliaires *avoir* et *être* est de pouvoir ainsi les distinguer des verbes pleins *avoir* et *être* lors du dépouillement des extrants.

Notre recherche portait sur une certaine évaluation non seulement de la richesse de la palette de temps verbaux employée par les répondants mais aussi de la qualité de cette utilisation. Tous les verbes ont donc été évalués et les formes qui sont jugées erronées sur le plan temps-mode-aspect se sont vu apposer un * après le code du temps. Les autres types d'erreur, tels les erreurs de genre, de personne, de nombre, de voix ou de pronominalization n'ont pas été codés à même le corpus.

Un deuxième code de temps est ajouté dans ces cas pour indiquer le temps que l'on aurait dû trouver à la place. Le codage *serait8*2* représente donc l'utilisation abusive du conditionnel à la

place de l'imparfait. Pour plus de facilité, nous appellerons dans l'analyse de cas similaires à celui mentionné ci-dessus «forme cible» la forme visée par le répondant mais non utilisée et «forme de substitution» la forme erronée qu'il a effectivement utilisée. Ainsi, nous pourrions dire que : «la forme de substitution au conditionnel se rencontre fréquemment à la place de la forme cible à l'imparfait dans des propositions conditionnelles introduites par *si*. Aux endroits où on est en droit de s'attendre à une forme cible autre qu'un verbe, la forme verbale de substitution est codée comme il se doit et est suivie de *, sans codage de forme cible (p. ex. : ...*le fini5** à la place de *la fin*, voir a.13.1.20); dans le cas inverse, où un vocable autre qu'un verbe est rencontré à la place d'un verbe, le vocable est suivi d'une * et du code de temps qu'aurait dû prendre la forme cible (p. ex. : ...*d'habit*9* à la place de ...*d'habiller*, voir a.22.5.44).

Les corpus informatisés et les extrants

Nous avons tenu à conserver une grande lisibilité aux corpus publiés dans le second volume. Afin de pouvoir effectuer des retours ponctuels à des endroits précis, nous avons indiqué en entête de chaque page le groupe linguistique et le numéro du sujet. En marge gauche du texte, nous avons, en champ fixe, fait mention du numéro de la page et de la ligne. Il est à noter que les lignes blanches portent le même numéro de ligne que la ligne les précédant. Dans les articles de ce présent volume, les renvois à des cas précis dans les corpus sont réalisés en faisant référence au groupe linguistique, au sujet, à la page et à la ligne. Les groupes linguistiques spécifiques sont représentés par les sigles suivants :

anglophone ..... «a»
francophone ..... «f»
lusophone ..... «p»

Ainsi la référence «a..4..1.34» doit être interprétée comme «sujet anglophone, numéro 4, page 1, ligne 34». C'est le même système de référence qui est en vigueur dans la présentation des extrants produits par JEUDEMO. Les chercheurs désireux d'exploiter les trois corpus informatisés ainsi que les extrants tels les listes alphabétiques, les listes de fréquences et les concordances peuvent en obtenir copie sur papier ou sur support

informatique à la demande (voir *Renseignements pratiques* à la fin du volume II).

Les listes alphabétiques[8] font mention des mots du texte, triés en ordre alphabétique normal ou inverse (mot lu de droite à gauche) et indiquent aussi accessoirement la fréquence du mot et les adresses où on peut les retrouver dans les corpus.

Les listes de fréquences[9] présentent les mots du texte, triés par ordre de fréquence, soit ascendante, soit descendante et relèvent également les adresses où ces mots sont situés dans les corpus.

Les concordances KWIC[10] (Key Word In Context) listent les mots concordés en milieu de page, entourés à gauche et à droite d'un bout de contexte, tel que trouvé dans les corpus. De plus, ils reprennent les adresses des mots concordés tout à gauche des extrants.

Quelques résultats

JEUDEMO nous fournit facilement des statistiques brutes sur les corpus. Il nous semble intéressant de faire des recoupements entre les résultats présentés ici et ceux qui émanent des divers articles de ce volume. À notre sens, le rapport entre la quantité de mots prononcés par l'interviewer et ceux prononcés par les sujets peut donner une indication de la loquacité des sujets. Une faible proportion de mots venant de l'interviewer peut vraisemblablement être attribuée au fait que le sujet est assez loquace et qu'en plus il ne demande pas à l'interviewer de reformuler les questions qu'il ne comprend pas. Le ratio est calculé en divisant le nombre de mots du sujet par le nombre de mots de l'interviewer. Comme les trois corpus informatisés n'ont pas été rassemblés selon les mêmes procédures et dans les mêmes circonstances, ni par les mêmes équipes, il est sans doute plus hasardeux de risquer des comparaisons entre les moyennes des différents corpus. Par contre, la différence s'estompe pour ce qui est d'une éventuelle comparaison entre les anglophones et les francophones Balablok, puisque *mutatis mutandis*, ils ont subi des protocoles d'entrevue similaires. Par ailleurs les comparaisons intra-corpus ne posent pas de problèmes. On peut aussi confronter les résultats individuels avec le ratio moyen calculé pour l'entièreté du corpus afin de situer les

performances des sujets les unes par rapport aux autres. Les trois tableaux ci-après donnent les résultats respectifs pour les anglophones, francophones Balablok et les lusophones ainsi que pour les corpus complets.

Nous pouvons constater à la lecture des résultats que les sujets parlant le moins ont connu des interventions de l'interviewer inversement proportionnelles aux leurs. Comme on pouvait s'y attendre, le groupe francophone Balablok a requis en moyenne des interventions légèrement moins nombreuses que le groupe correspondant anglophone. Les lusophones, il est vrai, placés dans une autre situation d'entrevue, ont nécessité en moyenne beaucoup moins d'interventions de la part des interviewers que les deux autres groupes.

CORPUS ANGLOPHONES

SUJET	NB. TOTAL DE MOTS	MOTS DU SUJET	MOTS DE INTERVIEWER	RATIO
01	1647	1385	262	18,91%
02	2304	1977	327	16,54%
03	2329	2071	258	21,55%
04	1537	1197	340	28,40%
05	1693	1304	389	29,83%
06	1042	814	228	28,00%
07	1475	1139	336	29,49%
08	2023	1748	275	15,73%
09	1838	1411	427	30,26%
10	2096	1761	335	19,02%
11	2594	2170	424	19,53%
12	1342	1067	275	25,77%
13	1481	1169	312	26,68%
14	1821	1630	191	11,71%
15	2074	1699	375	22,07%
16	1700	1444	256	17,72%
17	1905	1678	227	13,52%
18	1306	1090	216	19,81%
19	976	860	116	13,48%
20	1679	1466	213	14,52%
21	1310	948	362	38,18%
22	2639	2400	239	9,95%
23	2124	1684	420	24,94%
TOTAL	40915	34112	6803	19,94%

CORPUS FRANCOPHONES

SUJET	NB. TOTAL DE MOTS	MOTS DU SUJET	MOTS DE INTERVIEWER	RATIO
01	3132	2621	511	19,49%
02	2211	1705	506	29,67%
03	2406	1937	469	24,21%
04	2369	2087	282	13,51%
05	3251	2861	390	13,63%
06	1744	1401	343	24,48%
07	3019	2613	406	15,53%
08	2185	1934	251	12,97%
09	4083	3634	449	12,35%
10	1462	1124	338	30,07%
11	2592	2319	273	11,77%
TOTAL	28454	24236	4218	17,40%

CORPUS LUSOPHONES

SUJET	NB. TOTAL DE MOTS	MOTS DU SUJET	MOTS DE INTERVIEWER	RATIO
01	1255	1106	149	13,47%
02	1018	921	97	10,53%
03	821	702	119	16,95%
04	1598	1470	128	8,70%
05	3469	3327	142	4,26%
06	2356	2200	156	7,09%
07	2226	1980	246	12,42%
08	2041	1897	144	7,59%
09	2007	1666	341	20,46%
10	1165	721	444	61,58%
11	1887	1621	266	16,40%
12	1509	1308	201	15,36%
13	2120	1867	253	13,55%
14	2602	2375	227	9,55%
15	3183	3011	172	5,71%
16	1093	745	348	46,71%
17	1579	1293	286	22,11%
18	6792	6289	503	7,99%
19	3425	3113	312	10,02%
20	4717	4291	426	9,92%
TOTAL	46863	41903	4960	11,83%

Nous pouvons aussi distinguer dans quelle proportion les locuteurs ont recours à des mots de leur langue maternelle ou de la langue anglaise afin d'exprimer le plus clairement possible leurs idées. Pour ce faire, nous relevons les mots que nous avons entourés de barres obliques et qui sont répertoriés à la fin de chacun des JEUDEMO. Le ratio langue étrangère/français donne vraisemblablement une indication de la qualité de la connaissance du français et de la volonté d'apprendre une nouvelle langue le plus parfaitement possible. Une comparaison des résultats individuels avec les moyennes par corpus permet de situer les locuteurs les uns par rapport aux autres. Une comparaison inter-corpus donne des indications sociolinguistiques intéressantes. L'écart entre les trois groupes n'est pas très marqué et les différences sont plus sensibles entre sujets d'un même groupe pour ce qui est de l'utilisation de mots prononcés de façon non-standard ainsi que de mots de leur langue maternelle ou de l'anglais en ce qui concerne les groupes lusophone et francophone. Une étude attentive des différents corpus, lors de laquelle on distingue les mots prononcés de façon non-standard des mots n'appartenant pas au français, nous indique que les lusophones ont le moins tendance à utiliser des mots de portugais ou des mots de la langue anglaise quand ils en ignorent le correspondant français. Les lusophones semblent généralement vouloir reformuler les phrases de façon différente afin de contourner leur manque de vocabulaire. Les francophones ont recours à des termes anglais plutôt que d'essayer de reformuler leur intervention. Les anglophones passent du français à l'anglais dès qu'ils ne possèdent pas le vocabulaire nécessaire pour s'exprimer. Néanmoins, les diverses façons de pallier à un manque de vocabulaire changent d'un sujet à l'autre dans un même groupe de locuteurs. Ce ne sont que des tendances générales que nous avons voulu indiquer dans nos observations ci-dessus. Les trois tableaux ci-après rassemblent ces données; notons toutefois que les chiffres repris dans les colonnes «mots de LM ou A» recensent tous les mots codés entre barres obliques. Ils comptabilisent ainsi les mots de la langue maternelle ou de l'anglais ainsi que les mots à prononciation non-standard. Seule une lecture des extrants complets permet de faire le décompte des «mots de LM ou A» et d'arriver aux conclusions mentionnées ci-dessus :

CORPUS ANGLOPHONES

SUJET	MOTS DU SUJET	MOTS EN FRANÇAIS	MOTS DE LN OU A	RATIO
01	1385	1358	27	1,988%
02	1977	1951	26	1,332%
03	2071	2059	12	0,582%
04	1197	1172	25	2,133%
05	1304	1291	13	1,006%
06	814	812	2	0,246%
07	1139	1123	16	1,424%
08	1748	1739	9	0,517%
09	1411	1403	8	0,570%
10	1761	1735	26	1,498%
11	2170	2130	40	1,877%
12	1067	1066	1	0,093%
13	1169	1139	30	2,633%
14	1630	1612	18	1,116%
15	1699	1608	91	5,659%
16	1444	1441	3	0,208%
17	1678	1651	27	1,635%
18	1090	1060	30	2,830%
19	860	855	5	0,584%
20	1466	1443	23	1,593%
21	948	931	17	1,825%
22	2400	2366	34	1,437%
23	1684	1656	28	1,690%
TOTAL	34112	33601	512	1,523%

CORPUS FRANCOPHONES

SUJET	MOTS DU SUJET	MOTS EN FRANÇAIS	MOTS DE LM OU A	RATIO
01	2621	2573	48	1,865%
02	1705	1685	20	1,186%
03	1937	1923	14	0,728%
04	2087	2085	2	0,095%
05	2861	2821	40	1,417%
06	1401	1381	20	1,448%
07	2613	2578	35	1,357%
08	1934	1904	30	1,575%
09	3634	3519	115	3,267%
10	1124	1110	14	1,261%
11	2319	2306	13	0,563%
TOTAL	24236	23885	351	1,469%

CORPUS LUSOPHONES

SUJET	MOTS DU SUJET	MOTS EN FRANÇAIS	MOTS DE LM OU A	RATIO
01	1106	1031	75	7,274%
02	921	852	69	8,098%
03	702	694	8	1,152%
04	1470	1420	50	3,521%
05	3327	3307	20	0,604%
06	2200	2151	49	2,278%
07	1980	1934	46	2,378%
08	1897	1882	15	0,797%
09	1666	1642	24	1,461%
10	721	707	14	1,980%
11	1621	1609	12	0,745%
12	1308	1298	10	0,770%
13	1867	1843	24	1,302%
14	2375	2351	24	1,020%
15	3011	2984	27	0,904%
16	745	737	8	1,085%
17	1293	1288	5	0,388%
18	6289	6198	91	1,468%
19	3113	3082	31	1,005%
20	4291	4234	57	1,346%
TOTAL	41903	41240	663	1,607%

Notes sur la gestion informatique

Le soutien informatique d'un projet comme celui-ci comporte quatre aspects distincts : l'analyse, l'entrée/validation, la production et l'archivage. Les deux premiers points ont été évoqués plus haut; la production de listes en tous genres nécessite à chaque fois des adaptations de programmes afin de pouvoir les générer et, des fois même, l'ajout de codes dans les corpus pour faire ressortir plus facilement les traits linguistiques recherchés. Lors de chaque lecture ou chaque validation des fichiers quelques 830446 caractères étaient passés en revue. Ce total est supérieur à celui mentionné pour les trois corpus, car les fichiers informatiques contiennent en plus les caractères composant le champ fixe, les espaces, la ponctuation et les marqueurs de fin de ligne et de fin de paragraphe ainsi que les indications de changement de fonte de caractères. Sur une période de trois années, c'est très probablement plus d'un milliard de caractères qui ont été traités par les ordinateurs centraux du Centre de calcul. Afin d'éviter toute perte de fichiers suite à des défaillances matérielles, à des coupures d'électricité ou à des erreurs humaines, des copies de sécurité sont prises régulièrement sur ruban magnétique et sont conservées dans la voûte d'archivage.

Autres recherches et diffusion des résultats

Dans le seul domaine de l'apprentissage du français langue seconde, il y a très certainement encore un grand nombre d'études possibles sur les corpus présentés dans le deuxième volume. La structure syntaxique, l'enchaînement discursif, les champs sémantiques, pour n'en citer que quelques-uns, peuvent faire l'objet de recherches intéressantes. Les entrevues avec les lusophones seraient par exemple aussi une source précieuse d'information pour des études de sociologie de l'émigration.

Il entre donc dans l'intention des membres de l'équipe de recherche de rendre disponibles les corpus à la communauté des chercheurs. Il sera même possible d'en obtenir des copies sur support informatique en s'adressant au chercheur principal. Il sera également envisageable d'obtenir des listes générées par JEUDEMO portant sur les trois corpus ou sur des parties de ceux-ci. Des

recherches particulières nécessitant des listes contenant des mots précis peuvent aussi être commandées.

Notes et Références

(1) Les trois corpus informatisés représentent 363 pages de texte imprimé à simple interligne; le corpus anglophone en compte 137, le francophone BALABLOK en compte 93, alors que le lusophone couvre 132 pages. En termes de mots, les trois corpus représentent un total de 116232 mots qui se répartissent comme suit :

anglophones ..... 40915 mots
francophones ..... 28454 mots
lusophones ..... 46863 mots

Les trois corpus représentent un total de 489671 caractères dont la distribution entre corpus se ventile de la façon suivante :

anglophones ..... 174436 caractères
francophones ..... 114607 caractères
lusophones ..... 200628 caractères

La longueur moyenne des mots pour l'ensemble des corpus est de 4.08 caractères par mot et les différents corpus présentent les moyennes suivantes :

anglophones ..... 4.15 caractères/mot
francophones ..... 3.86 caractères/mot
lusophones ..... 4.12 caractères/mot

La longueur moyenne des formes différentes pour l'ensemble des corpus est de 7.29 caractères par forme et les différents corpus présentent les moyennes suivantes :

anglophones ..... 7.43 caractères/forme
francophones ..... 6.93 caractères/forme
lusophones ..... 7.40 caractères/forme

(2) Exemple de mise en page du corpus avec le champ fixe à gauche, suivi à la septième colonne du champ libre.

CORPUS ANGLOPHONE - SUJET 1
1..1 $ Tu pourrais nous reconter le film? %
1..1
1..2 c'est1 seulement - c'est1 l'animation - il
1..3 utilise1 pas les mots - c'est1 - il commence1 avec
1..4 des rectangles - qui sont1 comme des personnes et
1..5 ils marchent1 dans la ville et ils disent1 bonjour
1..6 à tous les autres rectangles et - maintenant il a1
1..7 un cercle qui marche1 dans la ville et les
1..8 rectangles s'arrêtent1 et /say/ oh - qu'est-ce qui
1..9 s'est passé3 - et ils - /follow/ - après - ils
1.10 suivent1 le cercle et lui battre9*1 - se battre9*1

(3) Dans l'exemple ci-dessus, la ligne 1..1 présente en champ libre, la première question du corpus anglophone, sujet 1. À son début se trouve le signe «$», l'embrayeur de langue «question» et après le point d'interrogation en bout de ligne, se trouve le signe «%», le marqueur de début de la langue «réponse». Les questions sont également soulignées et suivies d'un saut de ligne afin d'être facilement repérables.

(4) l'extrant de validation présenté en Annexe 1 montre une concordance servant à vérifier l'exactitude du codage des mots comportant une voyelle avec tréma. Dans nos textes et nos corpus le tréma est représenté par le signe = après la voyelle; lors de la sortie imprimée le signe est converti en tréma sur la voyelle. Par contre, les signes = sont codés <= dans les fichiers et ne sont transformés également que lors de l'impression.

(5) Ouellette Francine, (1983) *JEUDEMO*, Bulletin no. 98-01b, Montréal : Centre de Calcul de l'Université de Montréal.

(6) Un programme JEUDEMO présenté en Annexe 2 montre les trois parties : à savoir la description du texte traité, la description de la commande et enfin la description de la disposition des résultats lors de l'impression des extrants. Avant le début et après la fin du programme JEUDEMO en langage naturel se trouve une série de commandes en langage JCL (Job Control Language) afin de router le travail de manière satisfaisante dans tout l'énorme réseau informatique du Centre de Calcul.

(7) Voir l'instruction «Symb. alpha» dans l'extrant à l'Annexe 2.

(8) La liste alphabétique à l'Annexe 3 présente les mots triés en ordre alphabétique à gauche de la page, les adresses où on peut retrouver les mots dans les corpus et à l'extrême droite se trouve indiquée la fréquence de chacun des mots.

(9) La liste de fréquences à l'Annexe 4 donne à gauche des mots triés par fréquence décroissante l'adresse de la première occurrence de ce dernier et à sa droite est indiqué le nombre d'occurrences de ce mot dans le texte. À l'intérieur d'une même fréquence, les mots sont repris en ordre alphabétique.

(10) La concordance KWIC située à l'Annexe 5 présente les mots concordés en ordre alphabétique, disposés dans le milieu de l'extrant, avec à gauche et à droite de ceux-ci un court extrait du contexte entourant le mot dans le texte. À l'extrême gauche se trouve l'adresse du mot dans le texte.

Annexe 1

Concordance des = du Volume II - faite le 89/05/08 a 18.26.46. CENTRE DE CALCUL - sur le fichier f1

12080	work on the E.S.F. Project», in	Lu=di	, G. (éd.). _Devenir bilingue, parler
5720	ux études également (Raupach 1984,	Mo=hle	1984, Py 1984, Grandcolas et Soulé-Sus
11560	_ 35:1, 1-19. &Dechert, H.W., D.	Mo=hle	, et M. Raupach (éds). 1984. _Second
12010	au «TESL Canada Summer Forum». &	Mo=hle	, Dorothea. 1984. «A comparison of the
10900	d'une part, et l'évaluation globale, «	nai=ve	« ou «non nai=ve», d'autre part. Si l'on
10900	l'évaluation globale, «nai=ve» ou «non	nai=ve	«, d'autre part. Si l'on suppose que l'
11130	relation étroite entre l'évaluation «	nai=ve	«, d'un côté, et la mesure de variété
11160	incises, de l'autre; l'évaluation «non	nai=ve	« correspondrait plus étroitement aux
10830	même, en distinguant des évaluations «	nai=ves	«, subies par les anglophones et les
10840	les lusophones, des évaluations «non	nai=ves	«, qui avaient appliqué la procédure
11390	Spracherwerb ausla=ndischer Arbeiter_.	Tu=bingen	: Narr. &Connors, K., M-A. Vieira
11450	et de Philologie Romanes_, tome V,	Tu=bingen	: Niemeyer, pp. 307-19. &D'Anglejan,
11570	1984. _Second language productions_.	Tu=bingen	: Narr. &Deulofeu, J. et C. Noyau.
12090	_Devenir bilingue, parler bilingue_.	Tu=bingen	: Niemeyer. &Perdue, Clive (éd.).
5430	arabe espagnol finnois .not LC	<=	Langue cible .ligne LN <= Langue
5450	.not LC <= Langue cible .ligne LN	<=	Langue maternelle .text &Pour chaque

Annexe 2

```
texte.
jeudemo,s=toto.
kaki,toto,output.
exit,s.
kaki,toto,output.
*weor
********** ICI COMMENCE LE PROGRAMME JEUDEMO **********
********** DESCRIPTION DU TEXTE TRAITE **********
ligne          = 132
en-tete        = frequence1   89/05/08
symb.alpha     = a-à-â b c-ç d
.                e-é-è-ê f g h
.                i-ï j k l m n
.                o-ô p q r s t
.                u-ù-û v w x y z
.                tirettiret '' * /
.                0 1 2 3 4 5 6 7 8 9
symb.maj       = debut beng
symb.dia       = apres =
reference      = gauche 6
********** DESCRIPTION DE LA COMMANDE **********
commande       = liste conco longueur
********** DESCRIPTION DE LA DISPOSITION DES RESULTATS **********
disp.liste     = ref 1
contrainte     = freq 1-1
********** FIN DU PROGRAMME JEUDEMO **********
*weor
*pesc
*incl 1 * texte
*weor
```

Annexe 3

Liste alphabétique : langue QUESTION le 89/05/08 a 21.39.08. sur fichier rgb1 - page 1

a	a..1..2.39	a..1..4.17	a..1..5.34	a..1..6..1	a..1..6..7	
	a..1..6..9	a..1..6.16	a..1..6.29	a..1..6.29		9
à	a..1..2.28	a..1..3.13	a..1..6..1	a..1..6.18	a..1..6.19	5
après	a..1..3.31					1
armée	a..1..4.17					1
as	a..1..6.18	a..1..6.18				2
au	a..1..5..7					1
aurais	a..1..3.18	a..1..3.23				2
aurait	a..1..5.32					1
avaient	a..1..4.29	a..1..4.33				2
avait	a..1..4.11	a..1..4.35				2
avec	a..1..3..9	a..1..3.12				2
battent	a..1..4..6					1
battre	a..1..6.19					1
c	a..1..5.29	a..1..6.22				2
ça	a..1..2.23	a..1..3..9	a..1..3.12	a..1..3.17	a..1..3.26	5
canon	a..1..4.35					1
carré	a..1..3.23					1
carrés	a..1..3.12	a..1..5..7				2
ce	a..1..2.23	a..1..3.17	a..1..4.38			3
cercles	a..1..3..9	a..1..3.12	a..1..4.38	a..1..5..7		4
chefs	a..1..4.12					1
cherché	a..1..6..1					1
chose	a..1..4.35					1
combien	a..1..4.29	a..1..6.33				2
comme	a..1..3.17	a..1..3.26	a..1..4.35			3
commencé	a..1..2.39	a..1..4.11				2

Annexe 4

Liste frequence : langue QUESTION faite le 89/05/08 a 21.39.34. sur fichier rgbl - page 1

a..1..3..9	les	10	a..1..4.11	y	2	a..1..5..8	intelligence	1
a..1..2.39	a	9	a..1..3.31	après	1	a..1..6.25	interprétation	1
a..1..1..1	tu	9	a..1..4.17	armée	1	a..1..3.32	intolérants	1
a..1..2.28	à	6	a..1..5..7	au	1	a..1..5.29	jours	1
a..1..4.11	de	6	a..1..5.32	aurait	1	a..1..2.28	l	1
a..1..2.39	le	6	a..1..6..6	battent	1	a..1..3.23	m	1
a..1..3.10	un	6	a..1..6.19	battre	1	a..1..3.12	mais	1
a..1..2.23	ça	5	a..1..4.35	canon	1	a..1..3.23	moi	1
a..1..3.10	te	5	a..1..3.23	carré	1	a..1..6..9	négocié	1
a..1..3..9	cercles	4	a..1..4.12	chefs	1	a..1..6..7	négocier	1
a..1..2.23	est	4	a..1..6..1	cherché	1	a..1..5.29	nombre	1
a..1..2.34	et	4	a..1..4.35	chose	1	a..1..5.29	nos	1
a..1..1..1	le	4	a..1..4..6	comment	1	a..1..1..1	nous	1
a..1..2.23	ce	3	a..1..5..7	comparer	1	a..1..1.29	octogone	1
a..1..3.17	comme	3	a..1..5.34	condition	1	a..1..3.17	on	1
a..1..2.34	dans	3	a..1..3.31	d	1	a..1..4.11	ont	1
a..1..3.26	des	3	a..1..6.19	défendre	1	a..1..6.19	ou	1
a..1..4..6	ils	3	a..1..6.18	déjà	1	a..1..6.18	où	1
a..1..3.10	jeu	3	a..1..3.15	dée	1	a..1..6..1	paix	1
a..1..5..8	leur	3	a..1..6.22	difficile	1	a..1..4.29	parmi	1
a..1..2.23	qu	3	a..1..2.23	dire	1	a..1..2.34	pays	1
a..1..3..9	quand	3	a..1..4.33	en	1	a..1..3.13	penser	1
a..1..2.34	quel	3	a..1..5.29	encore	1	a..1..3.26	petits	1
a..1..2.39	qui	3	a..1..4.21	enrolée	1	a..1..5..8	point	1
a..1..2.28	une	3	a..1..2.28	époque	1	a..1..4.38	pour	1
a..1..6.18	se	2	a..1..6..7	essayé	1	a..1..1.12	pourquoi	1
a..1..3.18	aurais	2	a..1..3.23	étais	1	a..1..1..1	pourrais	1
a..1..4.29	avaient	2	a..1..6.18	été	1	a..1..5.32	pu	1
a..1..4.11	avait	2	a..1..5.32	être	1	a..1..2.29	quelconque	1
a..1..3..9	avec	2	a..1..3.31	êtres	1	a..1..5.34	quelle	1

Annexe 5

Concordance : langue QUESTION faite le 99/05/08 a 21.39.20. sur le fichier rgb1 - page 1

a..1..2.39	les personnes qui sont1 différentes $Qui	a		commencé le guerre?% au début c'est1 les
a..1..4.17	le chef $Dans une armée, tout le monde	a		le même importance?% dans une armée? - non -
a..1..5.34	$La guerre aurait pu être évitée?% oui $	A		quelle condition?% c'est1 difficile - si les
a..1..6..1	$Préjugés% préjugés $Lequel des deux	a		cherché à faire la paix?% les deux parce que
a..1..6..7	ont devient3 la même chose $Lequel des deux	a		essayé de négocier?% non - je pense1 pas
a..1..6..9	de négocier?% non - je pense1 pas $Qui	a		négocié?% ils juste battent1*3 et à la fin
a..1..6.16	mais je sais1 pas - on peut1 pas dire9 $Qui	a		gagné?% rien parce qu'ils sont1 la même
a..1..6.29	veux1 tu es1 peut--être différente que moi $	A		la fin, tout le monde a la même forme?% mais
a..1..6.29	différente que moi $A la fin, tout le monde	a		la même forme?% mais il y a1 une autre - non
a..1..2.28	arme et les mains $Tu peux situer l'histoire	à		une époque quelconque?% je ne sais1 pas - oh
a..1..3.13	cercles jouent avec les carrés ça fait penser	à		quel jeu?% oui oui le - /dice/ - $Les dés%
a..1..6..1	préjugés $Lequel des deux a cherché	à		faire la paix?% les deux parce que - les deux
a..1..6.18	as déjà été dans une situation où tu as eu	à		te défendre ou à te battre?% je peux1 pas -
a..1..6.19	une situation où tu as eu à te défendre ou	à		te battre?% je peux1 pas - je peux1 pas
a..1..6.22	je peux1 pas le raconter9 $C'est difficile	à		exprimer?% c'est1 difficile à exprimer9 -
a..1..3.31	sont1 de /polis/ - et ce n'est1 pas /fun/ $D'	après		toi, tous les êtres humains sont
a..1..4.17	et c'est1 lui qui est1*2 le chef $Dans une	armée		, tout le monde a la même importance?% dans
a..1..6.18	rien parce qu'ils sont1 la même $Tu as	déjà		été dans une situation où tu as eu à
a..1..6.18	$Tu as déjà été dans une situation où tu	as		eu à te défendre ou à te battre?% je peux1
a..1..5..7	$Tu peux comparer les cercles et les carrés	au		point de vue de leur intelligence et leur
a..1..3.18	on te taquinait comme ça, qu'est-ce que tu	aurais		fait?% première fois rien mais si ça
a..1..3.23	$Si tu étais un carré et moi un rond, tu m'	aurais		taquiné?% non $Des petits garçons
a..1..5.32	peu différence avec les fusées - $La guerre	aurait		pu être évitée?% oui $A quelle
a..1..4.29	les autres - quoi faire9 $Combien parmi eux	avaient		un fusil?% combien de quoi? $De fusils?% je
a..1..4.33	je comprends1 pas la question $Combien en	avaient		un fusil?% rien $Il n'y avait pas quelque
a..1..4.11	voir9 la - $Quand ils ont commencé, il n'y	avait		pas de chefs?% oui - avant que la guerre
a..1..4.35	$Combien en avaient un fusil?% rien $Il n'y	avait		pas quelque chose comme un canon?% je pense1
a..1..3..9	c'est1 difficile $Quand les rectangles jouent	avec		les cercles, ça te rappelle un jeu? Quel

Usage du système verbal par des apprenants du français langue seconde

Nives Bazergui

Dans ce chapitre, nous tenterons une étude de divers aspects du système verbal d'apprenants du français langue seconde. Nous croyons qu'une évaluation du système verbal est un indice très important du niveau de compétence atteint par ces locuteurs. Le premier volet de notre recherche s'intéresse surtout aux données fournies par l'encodage des formes verbales du corpus. Ainsi, ce sont les emplois du système de mode, temps et aspect du français que nous évaluons. Chaque forme erronée étant indiquée par un astérisque et accompagnée du code de sa forme cible, il devient aisé d'établir le pourcentage de formes correctes dans un corpus donné ou chez un individu. Des comparaisons entre groupes permettent de déterminer les domaines problèmes et les types d'erreurs particuliers ou généralisés. Des études ont été faites sur des aspects individuels du système verbal d'apprenants mais plus rares sont celles qui traitent du système verbal en général. Celui-ci pourrait-il être un indice valable de compétence? Est-il nécessaire d'en étudier des aspects plus significatifs tels l'emploi des auxiliaires principaux ou le système de constructions pronominales? Nous observerons donc ces deux derniers aspects de manière plus détaillée en comparant les données trouvées pour les apprenants à celles que nous ont procuré les deux corpus francophones étudiés pour cette recherche («Balablok» et «Thibault» - voir *Introduction*).

Ces deux corpus francophones nous serviront de norme d'acceptabilité surtout quant à l'emploi des auxiliaires.

Le codage du système verbal ne nous indique que le mode, le temps et l'aspect de la forme verbale. Nous étudierons donc cette dimension du système. Nous essaierons aussi de voir quelle influence peut avoir la langue maternelle de l'apprenant sur sa compétence du système verbal. Peut-on parler d'interférence? Dans le cas du corpus allophone, il ne nous sera pas possible de le déterminer.

Il est tout d'abord important de remarquer que nous parlerons de «formes verbales» (Connors et al. 1987) puisque Comrie (1985, pp.6-7) déplore, et avec raison, la tendance des grammairiens traditionnels à utiliser le terme «temps» pour désigner le temps et l'aspect. Une forme peut bien sûr différer d'une autre au niveau du temps et de l'aspect, mais, de plus, au niveau du mode puisque nous étudions aussi le conditionnel. Bien que le terme «forme verbale» soit trop général, nous nous y tiendrons pour ne pas entrer ici dans des problèmes théoriques et descriptifs trop profonds. Un terme tel «le temps présent» fera référence à ce qui est traditionnellement un sous-paradigme dont les formes peuvent différer d'autres sous-paradigmes au niveau de l'aspect et/ou du mode.

Le système mode, temps et aspect

L'observation la plus générale que nous pouvons faire des données est celle du pourcentage de formes grammaticales dans les corpus. La table suivante indique le nombre total de formes verbales dans les corpus (N), ainsi que le nombre d'erreurs (*N), le nombre de formes grammaticales (NG) et les pourcentages respectifs pour les trois groupes d'apprenants.

TABLE A POURCENTAGES DE FORMES GRAMMATICALES

	N	%	*N	%	NG	%
(N=23) ANGLOPHONES	6036	100	692	11,5	5394	88,5
(N=20) LUSOPHONES	7633	100	600	7,9	7033	92,1
(N=17) ALLOPHONES	4408	100	550	12,5	3858	87,5

Pour ces trois groupes, les écarts entre l'individu le plus fort et l'individu le plus faible sont indiqués dans la table suivante :

TABLE B ÉCARTS ENTRE INDIVIDUS

	+fort	+faible
ANGLOPHONES	97,2%	78,9%
LUSOPHONES	99,6%	55,3%
ALLOPHONES	94,2%	77,0%

% = pourcentage de formes correctes dans le corpus

C'est chez les lusophones que se manifeste la plus grande disparité entre les locuteurs. C'est parmi eux qu'on retrouve les plus faibles et les plus forts, ces derniers ne faisant presque pas d'erreurs au niveau du système verbal en général. Même si ces données ne l'indiquent pas, les anglophones forment un groupe beaucoup plus homogène que les deux autres. Quant aux allophones, il nous a semblé qu'une évaluation des formes verbales faite d'une façon aussi générale s'avérait peu révélatrice de leurs compétences puisqu'il était extrêmement difficile de juger de la grammaticalité de leurs formes sans prendre en considération, par exemple, la présence ou l'absence du sujet grammatical. En effet, plusieurs cas d'ellipse du sujet se manifestent dans le corpus allophone (170 cas, soit 4% des formes verbales) ce qui rend difficile la compréhension du discours. L'ellipse du sujet semble être un problème beaucoup plus généralisé chez les allophones que chez les deux autres groupes où seuls quelques rares cas ont été relevés.

Interférence «transparente» et confusion

Il s'agit donc de tenir compte de certains détails supplémentaires pour que les résultats changent ou pour que l'évaluation des divers groupes soit plus représentative. Pour comparer les anglophones et les lusophones, nous allons examiner quel rôle l'interférence peut avoir dans leur performance lors des entrevues suscitant l'emploi de formes verbales particulières. Nous nous attarderons sur nos données concernant le conditionnel, l'imparfait et le passé composé, ces trois formes verbales n'ayant pas le même champ d'usage en portugais, en anglais et en français. Le fait qu'il existe de fortes similitudes entre la typologie du français et celle du portugais sera sans doute la cause d'un certain degré d'interférence, ce qui ne devrait pas se manifester dans le discours des anglophones. Voici une table représentant les correspondances contextuelles partielles entre le portugais, l'anglais et le français quant au passé composé, à l'imparfait et au conditionnel du français.

TABLE C CORRESPONDANCES CONTEXTUELLES PARTIELLES

ANGLAIS	FRANÇAIS	PORTUGAIS
"present perfect"	passé composé	passé composé
"simple past"		passé simple
"past progressive"	imparfait	imparfait
conditionnel	conditionnel	

Les cas d'interférence «transparente» devraient se manifester dans les contextes suivants, différents selon qu'il s'agit de locuteurs anglophones ou lusophones : l'anglais standard parlé permettant l'emploi d'un conditionnel dans la position suivant le «si» d'une phrase hypothétique, nous nous attendons à ce que les anglophones emploient la forme verbale conditionnelle dans ce contexte en français (bien que cela puisse être dû à l'influence du français canadien non-standard); le portugais parlé utilisant le plus souvent l'imparfait dans la proposition indiquant le résultat de l'hypothèse, nous nous attendons donc, s'il y a interférence «transparente», à trouver une forte proportion d'imparfaits dans ces

propositions en français, calqués sur l'emploi portugais de l'imparfait de l'indicatif dans de telles propositions. L'imparfait de l'indicatif en portugais a une forme et un sens analogues à l'imparfait du français.

Nous étendrons le rôle de l'interférence à tout emploi du conditionnel au lieu de l'imparfait chez les anglophones et à tout emploi de l'imparfait au lieu du conditionnel chez les lusophones.

Pour ce qui est des narrations au passé, nous adopterons l'hypothèse de «confusion» pour les anglophones : les contextes pour l'imparfait et le passé composé chevauchent mais ne coïncident pas avec les contextes du "simple past" et du "present perfect". Le "simple past" empiète sur le territoire sémantique de l'imparfait et du passé composé. Nous nous attendons donc à trouver des cas d'imparfaits dans des contextes exigeant plutôt le passé composé et vice-versa. Des observations semblables s'avéreraient peu probables chez les lusophones comme le suggèrent les correspondances contextuelles de la table C. Plus généralement, toutefois, nous nous attendons à la «confusion» entre deux formes françaises (c.-à-d. le passé composé et l'imparfait, d'une part, et l'imparfait et le conditionnel, d'autre part), si et seulement si une seule forme en L1 correspond à chacune des deux formes françaises à la table C.

Observons maintenant les performances des deux groupes quant aux formes verbales traitées ici. La table D permet de voir que les lusophones ont moins de difficulté que les anglophones à employer cette partie du système verbal. Ceci est sans doute dû aux ressemblances générales entre le portugais et le français et, bien sûr, à la similitude de leurs systèmes verbaux.

TABLE D OCCURRENCES DES FORMES VERBALES

	ANGLOPHONES			LUSOPHONES		
	N	*N	*%	N	*N	*%
passé composé	653	88	13,5	820	53	6,5
imparfait	747	182	24,4	952	72	7,6
conditionnel	68	11	16,2	108	15	13,9

N = nombre total de formes verbales
*N = occurrences dont la forme remplace une forme cible de façon inadéquate
*% = *N/N

La table E indique que, pour les deux groupes, le passé composé semble plus facile à fournir là où il est requis, que l'imparfait. Si l'on compare ces données avec ce que l'on retrouve chez les allophones, on peut supposer que l'interférence ne joue pas un rôle important puisque l'on obtient des résultats comparables. Ces derniers semblent pourtant éprouver plus de difficulté avec l'imparfait et moins avec les deux autres formes que les anglophones et les lusophones.

TABLE E FORMES CIBLES

	ANGLOPHONES		LUSOPHONES		ALLOPHONES	
	N = 6036		N = 7633		N = 4408	
	*N	*%	*N	*%	*N	*%
passé composé	126	2,1	159	2,1	81	1,8
imparfait	174	2,9	216	2,8	178	4,0
conditionnel	73	1,2	112	1,5	27	0,6

N = nombre total de formes verbales du corpus
*N = occurrences verbales dont la forme remplace une forme cible de façon inadéquate
*% = *N/N

La constance avec laquelle, pour chaque groupe, le passé composé semble plus facile à fournir que l'imparfait, reflète les exigences différentes du français quant à ces deux formes.

L'hypothèse d'interférence «transparente» que nous avons proposée prédit que les anglophones remplaceraient l'imparfait par le conditionnel et que les lusophones remplaceraient le conditionnel par l'imparfait. Cela s'est-il produit? Les résultats sont présentés dans la table F.

TABLE F INTERFÉRENCE «TRANSPARENTE»

ANGLOPHONES		LUSOPHONES	
conditionnels dans les contextes exigeant l'imparfait :	1	imparfaits dans les contextes exigeant le conditionnel :	40
conditionnels choisis de façon erronée :	60	imparfaits choisis de façon erronée :	72

L'interférence «transparente» prédite pour les anglophones ne s'est pas produite puisque sur 60 choix erronés du conditionnel, seul un contexte exigeait l'imparfait. Les lusophones, par contre, laissent croire à la validité de l'hypothèse d'interférence puisque plus de la moitié (55,6%) des occurrences erronées de l'imparfait exigeait le conditionnel.

Quant aux hypothèses de «confusion», elles s'avèrent justes pour les anglophones, mais fausses pour les lusophones, comme le montre la table G. Les anglophones confondent l'imparfait avec le passé composé d'une part et avec le conditionnel de l'autre. Tel que prévu, ils ne confondent pas le passé composé et le conditionnel. Les lusophones, par contre, falsifient nos prédictions en confondant l'imparfait et le passé composé, ce qu'ils ne devraient pas être portés à faire puisque ces aspects du système temporel ne se chevauchent pas en français et en portugais. Tel que prévu, cependant, ils ont confondu l'imparfait et le conditionnel.

TABLE G «CONFUSION»

	ANGLOPHONES	LUSOPHONES
passé composé au lieu de l'imparfait :	18	*29
passé composé au lieu du conditionnel :	0	*2
imparfait au lieu du passé composé :	35	*31
imparfait au lieu du conditionnel :	16	40
conditionnel au lieu du passé composé :	0	0
conditionnel au lieu de l'imparfait :	1	10

* = non prédit par l'hypothèse de «confusion»

Le présent et l'infinitif à la place des formes cibles

En poursuivant notre observation de l'emploi du système verbal en général, il est intéressant de remarquer que les trois groupes ont tendance à souvent remplacer les diverses formes cibles par le présent et, dans une certaine mesure, l'infinitif, ce dernier phénomène se produisant surtout chez les allophones. Nous voyons ceci comme une abdication face aux complexités du système temporel, aspectuel et modal du français. La table H présente les données quant aux occurrences de formes du présent et de l'infinitif remplaçant diverses formes cibles.

TABLE H PRÉSENT À LA PLACE DES FORMES CIBLES

FORME CIBLE	ANGLOPHONES N	T	%	LUSOPHONES N	T	%	ALLOPHONES N	T	%
passé composé	58	126	46,0	104	159	65,4	32	81	39,5
imparfait	148	174	85,1	171	216	79,6	127	178	71,3
conditionnel	30	73	41,1	53	112	47,3	16	27	59,2

INFINITIF À LA PLACE DES FORMES CIBLES

FORME CIBLE	ANGLOPHONES N	T	%	LUSOPHONES N	T	%	ALLOPHONES N	T	%
présent	25	211	11,8	22	36	61,1	93	156	59,6
imparfait	5	174	2,8	0	216	0,0	22	178	12,4
passé composé	16	126	12,7	1	159	0,6	27	81	33,3
conditionnel	6	73	8,2	10	112	8,9	5	27	18,5

N = nombre total de formes au présent ou à l'infinitif remplaçant les formes cibles
T = nombre total de formes verbales remplaçant la forme cible
% = N/T

On remarque que pour les trois groupes, le présent tend surtout à remplacer l'imparfait. Le remplacement d'une forme cible par le présent semble donc être la solution principale adoptée par les apprenants pour surmonter la complexité du système mode, temps et aspect de leur langue seconde.

Quant à l'infinitif, il agit notamment chez les lusophones et surtout chez les allophones où il remplace souvent le présent et le passé composé. En effet, même chez les lusophones, 61,1% des formes erronées où le contexte exigeait le présent étaient des infinitifs. L'infinitif serait peut-être une porte de sortie supplémentaire surtout pour les allophones et, à un moindre degré, pour les anglophones et les lusophones.

L'évaluation du système mode, temps et aspect gagne à être complétée par l' évaluation d'autres aspects du système verbal du

français pour arriver à une image plus juste des apprenants. La prononciation, par exemple, étant un facteur important dans la distinction des formes verbales, il n'est pas toujours aisé de procéder à une évaluation précise en ne tenant compte que des critères considérés jusqu'ici.

Emploi des auxiliaires principaux

Parmi nombre d'autres critères, nous avons choisi, dans ce chapitre, de procéder à une observation de l'emploi des auxiliaires principaux (ÊTRE et AVOIR) chez nos trois groupes d'apprenants, en tenant compte de l'emploi de ces auxiliaires par les groupes contrôle francophones.

En effet, le français québécois diffère de la norme prescrite entre autres dans *Le bon usage*. Sankoff et Thibault (1977) en ont décrit l'usage à Montréal dans leur article traitant exclusivement de ces deux auxiliaires en français québécois. Cela nous servira de référence pour distinguer ce qui «sonne québécois» de ce qui est nettement une erreur, échappant à l'influence de l'environnement.

Ainsi, on constate que certains verbes sont plus aptes à être employés avec AVOIR que d'autres. Un facteur qui semble déterminant dans ce processus est l'aptitude du participe passé à servir d'adjectif exprimant un état, l'auxiliaire AVOIR ayant alors un rôle sémantique distinctif. Dans la conjugaison d'un verbe comme *changer*, par exemple, on comprendra pourquoi l'auxiliaire AVOIR est employé de façon si extensive puisqu'il permet une distinction aspectuelle. Exemple : *j'ai changé* vs. *je suis changé*. Cela est conforme à la norme prescrite.

L'emploi de l'auxiliaire AVOIR semble indiquer une *tendance vers la régularisation des conjugaisons* (Ibid., p.106), mais les formes les plus courantes y résistent plus longtemps. Certaines de nos données issues des corpus francophones montrent cependant que malgré cette résistance, même des verbes comme *venir* et *devenir* sont parfois conjugués avec AVOIR au Québec.

Canale, Mougeon et Beniak (1978) ainsi que Spilka (1976) ont étudié, entre autres, l'acquisition des auxiliaires principaux du français par des anglophones. Le degré d'exposition des appre-

nants au français québécois semble être un facteur important dans l'acquisition de ces auxiliaires. Nous avons observé trois groupes d'apprenants : anglophones, lusophones et allophones (asiatiques), qui ont eu des degrés différents d'exposition au français québécois. Leur apprentissage ne s'est pas fait dans les mêmes circonstances. Cela aura-t-il un effet sur leur compétence quant à l'emploi des auxiliaires?

L'évaluation générale du système verbal ne révélant pas de manière frappante le degré de compétence ou la ressemblance plus ou moins remarquable aux formes utilisées par un locuteur natif, nous pensons que le fait d'évaluer les locuteurs selon qu'ils «sonnent québécois» ou non pourrait être un indice significatif. Nous avons donc distingué les erreurs grammaticales des variations sociostylistiques dans notre évaluation de l'emploi des auxiliaires principaux.

Nos résultats se trouvent dans la table I qui suit. Nous avons relevé les formes verbales dont les auxiliaires ne suivaient pas les normes prescrites par les grammaires traditionnelles et nous nous sommes ensuite basés surtout sur les travaux de Sankoff et Thibault pour séparer des autres les emplois d'auxiliaires qui semblent tout à fait agrammaticaux, même en français québécois. Comme il s'agit d'un changement en cours, nous n'avons retenu que les cas extrêmes comme étant agrammaticaux. Nous donnons aussi le pourcentage des auxiliaires du corpus qui se classent dans cette catégorie.

TABLE I L'EMPLOI DES AUXILIAIRES

	N	*Aux	*%	Auxnq	%
ANGLOPHONES	578	45	7,8	18	40,0
LUSOPHONES	796	38	4,7	5	13,2
ALLOPHONES	442	33	7,4	12	36,4

N = nombre d'occurrences de formes composées (nécessitant un auxiliaire)
*Aux = nombre d'auxiliaires non conformes à la norme traditionnelle
*% = *aux/N
Auxnq = formes sociostylistiquement non-acceptables et non grammaticales
% = Auxnq/*Aux

Ces résultats permettent de voir un écart assez remarquable entre les groupes : la performance des lusophones laisse croire qu'en moyenne ils ont une compétence plus proche que les deux autres groupes de la compétence de francophones natifs quant à l'emploi des auxiliaires. On pourrait croire que le fait d'avoir vécu dans un contexte francophone plus longtemps que les allophones (qui pour la plupart avaient fini leur cours du COFI à peine six mois avant l'entrevue) et avec des contraintes plus fortes que les anglophones (le marché du travail, par exemple), a permis à ces lusophones d'acquérir un français plus ressemblant à celui que parlent les locuteurs natifs que les deux autres groupes. Comme Canale, Mougeon et Beniak (1978), nous pouvons en conclure que le degré de contact avec le français parlé à l'extérieur du contexte scolaire et la motivation, sans doute, jouent un rôle important dans l'acquisition d'éléments tels que les auxiliaires.

Il est à noter aussi qu'il existe chez les anglophones et les lusophones un lien étroit entre la performance au niveau des auxiliaires et leur succès dans leur usage du système de constructions pronominales (voir l'autre article de Bazergui dans ce volume).

Conclusion

L'étude du système mode, temps et aspect du français acquis en langue seconde ne semble pas aussi révélatrice que l'étude d'autres aspects du système verbal du français. Il semble que le mode, le temps et l'aspect soient si fondamentaux qu'en-deçà d'un seuil minimum de connaissance du système, toute conversation serait impossible. Ce pourcentage de formes grammaticales semble devoir se situer au-delà de 70%. Les différences entre les groupes ne sont pas frappantes à ce niveau-là. Il nous a donc semblé nécessaire d'ajouter d'autres dimensions à notre étude du système verbal : le rôle de l'interférence chez les deux groupes de langue indo-européenne, l'ellipse du sujet pour les allophones, l'emploi du système des auxiliaires pour les trois groupes. Ce dernier aspect nous a semblé particulièrement révélateur et nous mène à penser que l'ajout d'autres détails à une étude si générale est assez essentiel pour arriver à une évaluation plus exacte des groupes.

Références

Bazergui N., K. Connors, M. Lenoble et B. Majkrak. (1987). "Systematic variation in the use of the verb system by learners of French L2". Communication présentée au XIth University of Michigan Conference on Applied Linguistics.

Canale M., R. Mougeon et E. Beniak. (1978). "Acquisition of some grammatical elements in English and French by monolingual and bilingual Canadian students", dans *Revue canadienne des langues vivantes* 34 : 505-24.

Comrie B. (1985). *Tense*. Cambridge : Cambridge University Press.

Greviese M. (1986). *Le bon usage*, 12e édition refondue et corrigée par A. Goosse. Gembloux : Duculot.

Sankoff G. et P. Thibault. (1977). "L'alternance entre les auxiliaires AVOIR et ÊTRE en français parlé à Montréal", dans *Langue française* 34 : 81-108.

Spilka I. (1976). "Assessment of second-language performance", dans *La Revue canadienne des langues vivantes* 32 : 543-61.

Teyssier P. (1984). *Manuel de langue portugaise : Portugal-Brésil*. Paris : Klincksieck.

Acquisition et usage des constructions pronominales

Nives Bazergui

Cet article a pour objet l'emploi des constructions dites «pronominales» du français dans l'interlangue de locuteurs lusophones et anglophones, apprenants du français langue seconde. L'intérêt d'une telle analyse réside tout d'abord en l'ajout de détails supplémentaires à une étude plus générale de l'emploi du système verbal du français par ces mêmes locuteurs, tel que décrit dans l'article précédent de Bazergui. Le système de constructions pronominales (constructions en SE dès avant) étant particulier aux langues romanes, il est donc intéressant de pouvoir comparer deux groupes d'apprenants de L1 différentes selon ce critère. L'anglais ne fait pas partie des langues romanes, tandis que le portugais présente un système de constructions en SE semblable mais non identique à celui du français. Une comparaison de l'emploi du système en français par ces deux groupes d'apprenants permettra d'évaluer l'influence de la L1 dans l'apprentissage de cet aspect de la L2. Les nombreuses descriptions de ces constructions qui existent dans les diverses grammaires (Grevisse, 1986) et ouvrages théoriques (Boons, Guillet et Leclère, 1976; Wehrli, 1986) ne convergent pas toutes et il en ressort, avant tout, que l'on n'a pas encore de solution pour tous les problèmes posés par cet aspect de la grammaire. Plus rares sont les articles, à notre connaissance, qui traitent de l'acquisition de ces constructions en français L2 (Voir Spilka, 1976; Beniak, Mougeon et Côté, 1979; il s'agit surtout d'études faites auprès des enfants). On n'y propose aucune classification des constructions étudiées et on y emploie la terminologie «verbes pronominaux».

Nous tenterons ici d'exposer en bref les aspects les plus problématiques de la classification et de l'explication du fonctionnement des constructions en SE du français. Un même verbe pourrait, par exemple, être employé dans quatre constructions en SE, acquérant dans chacune d'elles une interprétation différente. De plus, le rôle du SE n'est pas le même dans chaque cas. Le même verbe peut aussi être employé dans des constructions sans le SE. Est-il alors légitime de parler de «verbe pronominal» quand c'est plutôt la construction dans laquelle il s'insère qui comporte certaines caractéristiques? Nous pensons qu'en français il existe un certain nombre de verbes qui ne peuvent pas s'insérer dans des constructions sans SE. Ce sont ces verbes seulement que nous qualifierons de «pronominaux», car ce sont ceux qui n'apparaissent jamais sans le SE clitique (p. ex. : *S'abstenir)*. Certains verbes, comme *apercevoir*, traditionnellement classifiés parmi les «verbes pronominaux» peuvent changer de sens selon la construction dans laquelle ils s'insèrent. Il nous semble donc préférable de ne pas les qualifier de «pronominaux».

Ils SE sont aperçus en traversant la rue. (interpr. réciproque)

Il S'aperçoit dans la vitrine du magasin. (interpr. réfléchie)

Il S'aperçoit de ses erreurs. (interpr. neutre)

La maison S'aperçoit du haut de la colline. (interpr. dite passive ou moyenne)

Les occurrences de SE + V doivent être étudiées individuellement et sont sujettes à l'un ou l'autre des types majeurs d'interprétation. C'est donc par l'interprétation sémantique d'une phrase surtout que peut se faire un classement des constructions en SE. Les ouvrages de linguistique traitant ce problème ne s'accordent pas tous quant aux tests syntaxiques permettant de caractériser ces constructions et de les classifier en groupes distincts. Ce qui est clair cependant, c'est que l'analyse porte sur toute une construction et non simplement sur le verbe, sur l'entrée lexicale. Il ne va donc pas de soi que l'on puisse dire d'un verbe qu'il est «pronominal à sens passif» quand c'est toute la construction dans laquelle il s'insère qui reçoit cette interprétation.

Hypothèses

En français et en portugais, les verbes sont sous-catégorisés comme pouvant s'insérer dans diverses constructions en SE. Cependant, tous les verbes ne peuvent s'insérer dans tous les types de constructions. En anglais, les verbes transitifs peuvent appeler un complément d'objet coréférent avec le sujet. Ce complément prend la forme d'un pronom réfléchi ou réciproque comme en français, à la différence qu'il est non-clitique. Les autres usages du SE français se réalisent syntaxiquement de façon différente en anglais. Il y a donc une correspondance très partielle entre cet aspect de la grammaire L1 des anglophones et leur grammaire L2. Nous prédisons que pour les cas clairs de SE pronom réfléchi et réciproque, les apprenants anglophones devraient éprouver moins de difficulté que pour les autres types de constructions, où le SE n'est pas clairement sémantiquement motivé. Le SE ne faisant partie de l'entrée lexicale d'un verbe que dans un nombre très limité de cas en français (Voir Boons, Guillet et Leclère, 1976), nous prédisons de la difficulté chez les anglophones à déterminer le rôle du SE dans les divers types de constructions. Peut-être aussi ne se rendent-ils pas compte de la productivité de certaines constructions. Leur L1, l'anglais, ne devrait avoir de l'influence qu'au niveau de certains choix lexicaux mais pas au niveau du fonctionnement du système puisque cette langue n'a pas de système comparable et que les fonctions assumées par les constructions en SE du français se traduisent par des constructions très diverses en anglais (Voir Enang, 1986). La difficulté pour ces apprenants viendra sûrement du fait qu'un même item assume tant de fonctions différentes[1].

Quant aux apprenants lusophones, nous prédisons une plus grande facilité d'emploi du système de constructions en SE que pour les anglophones, vu la présence d'un tel système dans leur L1. Celui-ci n'étant pas cependant identique à celui du français, certains usages sont peut-être plus largement permis en portugais qu'en français et vice-versa, et ce ne sont pas nécessairement les mêmes verbes qui sont sous-catégorisés comme pouvant s'inscrire dans les mêmes constructions pour les deux langues. Il y a donc possibilité d'un certain degré de confusion et d'interférence entre les systèmes L1 et L2 des lusophones.

Méthodologie

Notre étude consiste en l'analyse des corpus LUSOPHONE (N = 20) et (BALABLOK) ANGLOPHONE (N = 23) pour les données d'apprenants (voir en Introduction la description des corpus). Nous nous sommes aussi servie de deux groupes contrôles francophones étant donné les différences entre les entrevues menées auprès des apprenants. Le contenu de l'entrevue jouant pour les anglophones un rôle assez déterminant quant aux constructions en SE, nous n'avons pas pu comparer les deux groupes d'apprenants quantitativement sans passer par des groupes témoins francophones comparables aux lusophones d'une part et aux anglophones de l'autre. Nous avons donc consulté le corpus FRANCOPHONE «BALABLOK» (N = 11) comme contrôle pour les anglophones et FRANCOPHONE «THIBAULT» (N = 17) pour les lusophones (voir en Introduction la description des corpus).

Nous avons procédé à un dépouillement des quatre corpus en relevant les occurrences de constructions en SE ainsi que les cas d'ellipse du SE et autres formes erronées chez les apprenants.

Résultats

Voici tout d'abord un tableau des résultats quant au nombre d'occurrences des constructions en SE dans les quatre corpus consultés.

TABLE A NOMBRE D'OCCURRENCES

	BALABLOK FRANCO (N=11)	BALABLOK ANGLO (N=23)	FRANCO THIBAULT (N=17)	LUSOPHONES (N=20)
total V	3276	6036	3223	7641
total SE	490	278	153	289
% SE	15,0%	4,7%	4,7%	3,8%
moyenne par locuteur	45	12	9	15

% SE = total SE/total V

La proportion élevée de constructions en SE chez les FRANCOPHONES BALABLOK s'explique sans doute par la nature de l'entrevue qui portait sur un dessin animé décrivant l'interaction de deux groupes ennemis, incitant donc à l'emploi de nombreuses constructions à interprétation réciproque. Cela va de même pour les anglophones mais, comme nous le verrons plus loin, il semble qu'il s'agisse justement là d'une source de problèmes pour eux. Une comparaison des résultats des deux groupes d'apprenants par rapport aux groupes francophones témoins montre que les anglophones emploient beaucoup moins de constructions en SE que les Portugais qui sont proportionnellement assez près des FRANCOPHONES THIBAULT. On peut donc en conclure que les anglophones sous-utilisent le système de constructions en SE du français. Les taux des lusophones et des anglophones sont presque les mêmes malgré la nature très différente des entrevues.

TABLE B ELLIPSE DU SE

	BALABLOK ANGLO (N=23)	LUSOPHONES (N=20)
total SE	281	289
ellipses SE	73	9
% verbes	1,2%	0,1%

% verbes = ellipses/total V

On remarque que le phénomène de l'ellipse du SE existe chez les anglophones de manière beaucoup plus significative que chez les lusophones. C'est en fait le type majeur d'erreur trouvé chez les anglophones. Le nombre de cas d'ellipse dans les constructions contenant le verbe *battre* et les verbes du même champ sémantique (tels *batailler, chicaner, frapper)* est remarquable chez eux. Il est à noter cependant qu'il y a un taux élevé d'occurrence de ces verbes employés correctement. La table suivante indique les résultats quant à ces cas d'ellipses.

TABLE C ELLIPSES AVEC *BATTRE*

	BALABLOK ANGLO
total ellipses	73 (=100%)
ellipses battre	44 (= 60%)
même champ sém.	11 (= 15%)

% = ellipses battre/total ellipses

Les résultats révèlent que 75% des ellipses se produisent dans des constructions contenant le verbe *battre* ou des verbes du même champ sémantique. Le verbe *battre* peut s'insérer dans des constructions à interprétations sémantiques différentes. Les fonctions du SE dans ces constructions se révèlent dans les exemples avec *battre* comme n'étant pas fixées dans l'esprit des apprenants anglophones. Le verbe *battre* s'emploie avec un SE pronom sémantiquement motivé dans les cas où il s'agit d'une action réciproque ou réfléchie, exercée sur soi-même; mais dans un grand nombre de cas, le SE ne semble pas sémantiquement motivé dans la construction (ex. : *Il SE bat contre l'ennemi*). C'est sans doute ce continuum de fonctions sémantiques du SE pour un même verbe qui déroute les apprenants anglophones. Il y a inconstance marquée chez certains d'entre eux; ils semblent associer l'usage de certains verbes avec l'usage du SE sans cependant avoir assimilé les conditions d'insertion de ces verbes dans une construction en SE. La valeur, la fonction du SE ne semble donc pas être comprise, il n'est pas employé selon des critères clairs chez eux.

Chez les lusophones, par contre, l'ellipse ne semble pas constituer un problème majeur et, pour des raisons phonologiques, les quelques cas relevés ne sont pas toujours clairs. En voici quelques exemples :

p.18.1.14 *ils ? sont organisé une petite vie*

p.11.3.32 (ellipse du sujet) *quand ? sont mariés il y a 22 ans*

Il y a cependant chez eux, comme chez les anglophones, des cas de SE superflus. Un SE a été introduit dans une construction où il n'était pas requis ou, pour les lusophones, l'usage d'une construction a été étendu au-delà de l'usage habituel du français. Voici un tableau des résultats concernant le SE superflu chez les deux groupes.

TABLE D SE SUPERFLUS

	BALABLOK ANGLO	LUSOPHONES
SE superflus	32	21
% verbes	0,5%	0,2%

% verbes = SE superflus/total V

Le phénomène du SE superflu doit être étudié de façon différente pour les deux groupes, car il semble affecter des types différents de constructions chez chacun d'eux.

Pour les anglophones nous n'avons pas réussi à trouver de systématicité dans l'emploi superflu du SE. Un critère qui, selon nous, pourrait motiver certaines de ces erreurs est l'analogie sémantique. Ainsi, *X...*S'arriver* dans a.23.4.12 *quend le grand bataille S'arrive* ou a.17.2.33 *la même chose va S'arriver* pourrait être dû à une analogie faite avec, par exemple, *la même chose va SE passer*. *X...*SE marcher* pourrait être vu comme une analogie avec *X...SE promener*. *X...*SE devenir..Y* tirerait peut-être sa source dans *X...SE transformer en...Y* ou *X...SE changer en...Y*. En tenant compte de l'importance du phénomène de l'ellipse chez les anglophones montrant une difficulté à maîtriser le système, la présence de SE superflus semble pouvoir s'expliquer par le biais de l'analogie sémantique qui pousserait les apprenants à insérer un verbe dans une construction en SE parce qu'ils ont maîtrisé une construction synonyme en SE. Chez les lusophones, les cas de SE superflus ont deux causes principales : le lexique d'une part et la surextension d'une construction en français de l'autre. Sur 21 cas de SE superflus, 13 sont des cas de l'emploi de la construction dite «à sens passif» là où le français préfère une construction impersonnelle en ON. En voici deux exemples :

p.2.3.24 *ici *SE travaille beaucoup*

p.6.4.50 *après ça *SE met du poivre*

Teyssier (1984) remarque que «la forme pronominale comme équivalent du français ON (sens passif) est très courante en portugais» (p.197, note 4). Les autres cas de SE superflus peuvent sans doute s'expliquer par un transfert lexical du portugais, c'est à dire qu'un verbe est introduit dans une construction pour laquelle il n'est pas sous-catégorisé en français. Exemples :

p.11.1.27 *je *ME rencontre dans une...*

p.14.5..5 *le porc il *SE riait de l'âne*

p.20.4.29 *cet ami *S'encadre dans ma conception d'amitié*

Une dernière particularité de l'emploi du système, commune aux deux groupes d'apprenants bien que très peu répandue, est le mauvais placement du SE dans la phrase. En français, le SE se place en position clitique devant le verbe lexical ou son auxiliaire le cas échéant, mais non devant un semi-auxiliaire (comme cela peut être le cas en portugais (Teyssier, p.95)). Chez les lusophones, les trois cas relevés sont des constructions dont le SE de la forme cible s'insère normalement entre un verbe semi-auxiliaire conjugué et un infinitif.

p..5.1.44 *quelque chose qui *SE pouvait ▼ demander*

p..6.1.41 *il *SE vient ▼ marier avec une fille*

p.19.1.44 *je *ME commence à ▼ intégrer*

Quant aux anglophones, il s'agit pour la plupart d'un placement erroné du clitique entre un verbe et son auxiliaire, alors que le clitique en français doit se placer avant l'auxiliaire. Il est à noter aussi que l'auxiliaire *avoir* a été conservé.

a.21.3..3 *les cercles ▼ ont *SE mis dans une situation amusante*

a.22.1.14 *ils ▼ ont *SE moqué de...*

a.23.6.42 *j' ai seulement *ME protégé*

Finalement, un type d'erreur très peu fréquent mais attesté au moins une fois chez chacun des groupes consiste en l'assignation de la mauvaise personne au clitique (faute d'accord).

a.21.2.31 *je veux pas *SE battre*

p..6.4.52 *Attends ! *SE trompe!* (pour je ME trompe)

Le phénomène pourrait être assez révélateur du fonctionnement du système à savoir si le SE fait partie de l'entrée lexicale, s'il est acquis comme partie intégrante du verbe plutôt que comme pronom qui prend la marque de la personne. Il y a cependant bien trop peu d'exemples pour pouvoir en tirer quelque conclusion que ce soit. Un test visant à cerner ce phénomène serait sans doute intéressant.

Finalement, citons une particularité relevée chez les anglophones, assez significative selon nous puisqu'elle cerne les constructions à interprétation réciproque (ceci s'applique dans quelques cas aux réfléchies). Ce type de construction est reconnu dans toutes les descriptions grammaticales des constructions en SE. Il semble y avoir une tendance chez les anglophones à l'éviter, comme le montrent les exemples suivants :

a..1.1.36 *toutes les personnes marchent dans la ville et dit bonjour bonjour*

a..3.1..5 *là ils jettent les choses un à l'autre*

a..8.1..5-6 *si ils parlaient ensemble*

a.22.2..8-9 *il* (pl) *peut être la même chose, il peut ressemble à la même chose*

a.22.5.17-8 *tu ne peux pas aider toi-même, tu dois connaître toi-même*

a.16.5.39 *je défends pas moi-même*

a.11.8.20 *on regarde nous sur le film*

La réalité grammaticale du SE ne semble pas être comprise puisque, pour certains, même là où il est nettement pronom, dans les cas les plus clairs il y a confusion dans l'usage. Cela est indiqué par l'évitement du réciproque. Le SE ne semble donc pas représenter une réalité réciproque ou réfléchie.

Conclusion

L'étude de l'emploi de constructions en SE du français par des apprenants lusophones et anglophones révèle une nette influence de la L1 dans le taux de succès en production. En effet, le groupe ayant une L1 possédant un tel système semble avoir plus de facilité à le manier que celui pour qui l'emploi du SE est une particularité de l'interlangue. Il ne faut cependant pas se limiter à une observation de corpus pour tirer de telles conclusions. Une étude portant spécifiquement sur cet aspect de la grammaire, en insistant sur des conditions identiques pour les deux groupes étudiés serait de rigueur.

Nous pensons cependant que les Portugais continueront à employer les constructions en SE à sens passif en français jusqu'à preuve du contraire puisque le français utilise cette même construction de façon productive mais dans moins de contextes que le portugais. Quant aux anglophones, l'analyse du corpus ne permet pas une évaluation satisfaisante de la réalité psychologique du SE pour eux. On constate cependant par le nombre et le type d'erreurs que ce système n'est pas stable dans leur grammaire d'interlangue.

Cet aspect de la grammaire du français pourrait s'avérer assez intéressant dans le cadre d'une théorie de l'acquisition des langues secondes puisqu'il relève de plusieurs niveaux d'analyse grammaticale à la fois, soit le lexique, la syntaxe et la sémantique. La fonction et la nature du SE restent difficiles à définir pour les grammairiens, que peuvent en faire les apprenants?

Note

1. Nous n'entrerons pas ici dans la discussion à savoir si le SE est un élément polysémique ou s'il s'agit de plusieurs SE homonymes, mais le débat est présent dans la littérature (voir entre autres Wehrli [1986]).

Références

Beniak E., R. Mougeon et N. Côté. (1979). "Acquisition of French pronominal verbs by groups of young monolingual and bilingual Canadian students", dans *The Sixth LACUS Forum*. W.C. McCormack et H.J. Izzo (éds). Columbia, S.C., Hornbeam Press : 355-68.

Boons J.-P., A. Guillet et C. Leclère. (1976). *La structure des phrases simples en français : constructions intransitives*. Genève, Droz.

Enang A. (1986). *The translation of the French reflexive verbs into English*. Mémoire de Maîtrise. Université de Montréal.

Grevisse M. (1986). *Le bon usage*. 12e édition refondue et corrigée par A. Goosse. Gembloux, Duculot.

Spilka I. (1976). "Assessment of second-language performance", dans *La Revue canadienne des langues vivantes* 32 : 543-61.

Teyssier P. (1984). *Manuel de langue portugaise*. Paris, Klincksieck.

Wehrli E. (1986). "On some properties of French clitic SE", dans *Syntax and Semantics 19 : the Syntax of Pronominal Clitics*. H. Borer (éd.). New York, Academic Press : 263-83.

Quelques aspects de la cohésion textuelle : formes pronominales

Béatrice Majkrak

L'étude des corpus des apprenants de langues secondes peut être abordée de plusieurs façons différentes en fonction du but poursuivi : notre objectif étant de trouver des indices de la compétence de communication des locuteurs du français langue seconde, nous nous sommes concentrée sur la notion de la compréhensibilité du message. Nous voulons réaliser une étude comparative portant sur le comportement linguistique face aux renseignements à transmettre. La notion qui sous-tend une telle étude est celle de la cohésion textuelle, c.-à-d. de la stabilité du texte à travers la continuité des occurrences d'unités porteuses de sens (R. de Beaugrande, W. Dressler, 1972 : 36 et 48). Il ne s'agit pas de la description des modèles obéissant aux règles explicites de la grammaire : en réalité nous essayons tout d'abord d'examiner l'influence de certains facteurs cognitifs sur l'emploi de la syntaxe, c.-à-d. d'étudier par exemple les conséquences des limites de la mémoire à court terme dans les conditions d'entrevue sur le traitement linguistique des renseignements à transmettre.

Pour pouvoir suivre le discours d'une personne, il nous est indispensable de savoir à quoi se réfèrent les termes composant le message. Ceci est lié tant à des phénomènes extralinguistiques (connaissance d'une réalité, d'un domaine, d'un fait particulier, etc.) qu'à la manière dont le locuteur organise ses propos. Autrement dit, la compréhensibilité du message dépend autant de sa cohérence que de sa cohésion. Quand, par exemple, le locuteur

choisit de remplacer plusieurs termes apparaissant dans un court segment du discours, et qui renvoient à des référents différents, par un ou plusieurs pronoms, il risque de diminuer considérablement (dans certains contextes) le caractère explicite de son message. Une forme pronominale renvoyant à un terme excessivement éloigné dans le texte produit le même effet. Le phénomène contraire, c.-à-d. l'emploi répétitif des mêmes termes à courte distance augmente la redondance, étant le signe tantôt de l'intention particulière du locuteur (mise en évidence du terme, insistance persuasive, but stylistique, etc.), tantôt de son insécurité linguistique (ignorance ou peu de connaissance des règles de placement de clitiques, etc.). L'emploi des formes pronominales se situe ainsi à mi-chemin entre la pragmatique et la connaissance technique du système linguistique (B. Lust, 1986 : 42-3, 82-3). Leur étude permet de juger du degré d'expérience et d'assurance linguistique du locuteur à la condition de procéder à une comparaison systématique avec le comportement normal (habituel) dans ce domaine du locuteur natif.

Au début de la présente recherche nous nous sommes ainsi posée les questions suivantes : Comment et avec quelle fréquence les mots ou les structures linguistiques déjà employés sont-ils répétés, modifiés ou simplement omis (sous supposition que la répétition serait redondante)? Quels sont les modèles syntaxiques le plus souvent impliqués dans de telles opérations? Quelle est l'étendue maximale du texte contenant plus ou moins explicitement des mots ou des expressions renvoyant à un même référent? Les caractéristiques de l'équilibre entre l'économie d'expression et le degré de clarté sont-elles les mêmes dans les corpus des apprenants de langues secondes et ceux des locuteurs natifs?

Notre étude ne cherche pas à trouver ou à établir un modèle normatif du traitement linguistique des référents : nous avons en réalité décidé de considérer comme norme l'ensemble de nos corpus témoins sans cependant l'utiliser pour l'interprétation de tout écart comme erreur. Nous nous sommes attendue plutôt à ce que les différences trouvées puissent mettre en évidence des indices de la réussite ou de la compétence de communication des apprenants du français langue seconde.

Les textes auxquels nous avons affaire sont destinés à être écoutés : il s'agit d'entrevues enregistrées et transcrites par la suite. Le protocole d'entrevue n'a pas cependant été le même pour tous les quatre ensembles de corpus. En réalité, un protocole identique a été employé seulement pour deux ensembles de corpus : les corpus anglophones Balablok et les corpus francophones Balablok (des corpus témoins). Les corpus des locuteurs lusophones et les corpus Thibault 84 se situent très près les uns des autres par le type d'entrevue, tout en étant constitués en fonction d'objectifs différents et par des équipes de chercheurs non reliées. Tandis que les deux premiers groupes de locuteurs décrivent - en guise de réponse à une liste précise de questions - le contenu d'un film à court métrage (dessin animé), les deux autres parlent de leur vie quotidienne (famille, vie de quartier, travail) en répondant à des questions dont le nombre et le contenu peuvent varier en fonction de la loquacité du locuteur ou de sa volonté de parler de certains sujets, etc. La longueur des corpus étant souvent assez disparate, la section de l'entrevue retenue pour l'analyse ne comprend pas nécessairement les réponses aux questions identiques. Dans un souci d'uniformité des conditions d'analyse, nous avons toutefois retenu pour tous les corpus une longueur de texte correspondant à environ cinq pages dactylographiées de texte suivi par locuteur (100 lignes de production orale nette par locuteur à l'exclusion des questions et des interventions de la part des interviewers).

Quoique consciente dès le début des différences fondamentales entre les deux types de protocoles employés, ce n'est qu'après une analyse préliminaire de quelques échantillons que nous avons constaté la nécessité de mener notre analyse séparément pour les deux sous-ensembles de corpus ainsi différenciés. Or, il est devenu évident que le nombre d'occurrences de certains phénomènes étudiés était trop peu significatif dans le cas des corpus faits d'après le protocole no 1, dans d'autres cas ceci était vrai pour les corpus faits d'après les protocoles nos 2 et 3. Nous avons alors décidé de traiter séparément les deux groupes de corpus Balablok d'une part et l'ensemble des corpus Thibault ainsi que les corpus des locuteurs lusophones d'autre part, en privilégiant pour chacun des deux sous-ensembles l'étude de phénomènes différents.

Pour les deux ensembles de corpus Balablok, c'est l'étude des formes pronominales qui a été au centre de notre attention. Ayant pour tâche de raconter le contenu d'un film (déroulement d'une guerre entre deux groupes de personnages géométriques se transformant graduellement pendant les combats en une autre forme géométrique), les locuteurs ont dû se servir - de façon assez prévisible - d'un certain nombre de termes (personnages, outils, descriptions du lieu de l'action, etc.) qui revenaient très fréquemment au cours de toute l'entrevue. Très souvent, deux possibilités en découlaient : la répétition du terme ou sa substitution par une forme pronominale. Nous avons pu ainsi observer quelle est la fréquence de substitutions de termes par des formes pronominales, quelle est la distance moyenne et maximale entre le terme remplacé et la forme pronominale correspondante, s'il y en a, et quelles sont les fonctions grammaticales qui semblent favoriser l'apparition des formes pronominales.

Méthodologie

La recherche s'est déroulée en plusieurs étapes :

Dans un premier temps, nous avons préparé tous les corpus en délimitant les unités d'analyse de base. Travaillant avec des entrevues enregistrées, nous avons défini comme unité de base une étendue de texte comprise souvent entre deux blancs prosodiques (pauses plus ou moins longues signalant la fin du propos). Les segments ainsi obtenus ne correspondent souvent pas à des phrases au sens traditionnel (vu le nombre élevé de propos inachevés, d'hésitations, de changements brusques de structures, etc.), mais ce sont plutôt des unités cohérentes au sens d'énoncé (unité logique).

> /pour trouver un travail j'ai - dit-on - j'ai été pas trop pressé pour ça/ c'était pas en rush là ça parce que j'étais chez ma soeur/

Ensuite, chacun des corpus témoins et des corpus des apprenants du français langue seconde a été analysé en vue du repérage des substitutions de termes. Ainsi, dans chaque corpus nous avons relevé toutes les formes pronominales en tenant compte de leur coréférents (les termes remplacés).

/DEUX PETITS CARRÉS voient UN ROND/ puis là ILS jouent à balle avec LUI/

Par après, toutes les répétitions de termes à sens explicite à courte distance (trois unités de base ou moins, y compris à l'intérieur de la même unité de base) ont été relevées.

/LES BALLES étaient maintenant aussi fortes que LES BLOCS/ LES BALLES ont attaqué LES BLOCS/

Pour chacune des formes étudiées, la fonction grammaticale a été déterminée. En même temps, nous avons tenu compte de la fonction grammaticale du coréférent. Après avoir effectué le relevé pour un échantillon représentatif des corpus étudiés, nous avons abouti à quatre catégories possibles, dont chacune a reçu un code (1-4). Pour des raisons pratiques, nous avons cependant décidé d'ajouter une autre catégorie (5) pour les formes «y», «en» (remplaçant un terme ou une partie plus ou moins longue de l'énoncé ou l'énoncé au complet), ainsi que «le» remplaçant toute une proposition. En résumé :

Catégorie 1 : coréférent remplissant la fonction du sujet remplacé par une forme pronominale à la même fonction

/LE CARRÉ se rend compte que.../ IL (1) veut expliquer ça à tout le monde/

Catégorie 2 : coréférent sujet - forme pronominale objet

/LES CARRÉS étaient beaucoup/ je LES (2) ai pas compté.../

Catégorie 3 : coréférent objet - forme pronominale sujet

/le carré a vu UN ROND/ IL (3) était un peu drôle/

Catégorie 4 : coréférent objet - forme pronominale objet

/il voit UN BONHOMME par terre/ il vient pour LUI (4) donner un coup/

Catégorie 5 : y, en, le

/je sais pas ça s'appelle là L'AFFAIRE/ il y EN (5) avait assez dans l'ancien temps/

Tous les cas d'erreurs dans la réalisation des formes pronominales ont été relevés.

/je suis arrivée pour LUI battre/

Les cas de la présence dans la même unité de base de deux formes pronominales ayant des coréférents différents ont été relevés.

/le rond s'approche du carré/IL LE regarde/

Toutes les occurrences ont été relevées en termes d'éloignement entre le coréférent et la forme pronominale correspondante en fonction des unités de base :

(0) : à l'intérieur d'une même unité de base

/LA BALLE avait peur et ELLE courait/

(1) : après un séparateur d'unité de base

/ils sont partis à courir après LES CARRÉS/ ils LES ont enfargés/

(2) : après deux séparateurs d'unité de base

/ils sont partis à courir après LES CARRÉS/ ils les ont enfargés/ ils LES ont battus/

(3) : après trois ou plus de séparateurs d'unité de base

/il arrive pour taper sur LE CARRÉ/ puis comme il arrivait pour taper il s'est aperçu de quelque chose/ il a arrêté/ il L'a regardé .../

Compilation et analyse des données

Les résultats compilés dans les tableaux ci-joints mettent en évidence des différences importantes existant à plusieurs niveaux entre le traitement linguistique des référents chez les locuteurs natifs et chez les apprenants du français langue seconde.

TABLEAU 1 BALABLOK FRANCOPHONES

	Type de forme pronominale								
SUJET	1	2	3	4	5	E	TOTAL	2FP	R
1	54	3	14	1	4	13	89	4	7
2	58	4	3	3	2	10	80	8	2
3	30	5	-	5	5	3	48	3	2
4	65	8	14	4	3	2	96	14	2
5	94	7	10	7	11	10	139	9	4
6	32	2	6	2	3	15	60	1	4
7	67	6	2	6	4	6	91	7	3
8	29	5	4	1	6	6	51	3	4
9	36	4	1	2	-	6	49	3	3
10	26	2	6	-	6	5	45	2	1
11	85	1	4	-	9	7	106	3	5
TOTAL	576	47	64	31	53	83	854	57	37

	Distance (type no 1)				
SUJET	0	1	2	3+	?
1	7	10	10	27	1
2	7	10	4	37	1
3	5	10	3	12	-
4	2	17	9	37	-
5	16	35	8	35	-
6	2	5	3	22	4
7	8	7	4	48	1
8	5	9	5	10	1
9	1	13	7	15	-
10	6	11	5	4	1
11	11	30	17	27	-
TOTAL	70	157	75	274	9

2FP = 2 formes pronominales différentes dans 1 propos
E = reprise du terme emphatique
? = formes pronominales à coréférent inconnu
R = répétition du même terme

TABLEAU 2 BALABLOK ANGLOPHONES

	Type de forme pronominale								
SUJET	1(*)	2(*)	3(*)	4(*)	5(*)	E(*)	TOTAL(*)	2FP(*)	R
1	27	-	10	2(2)	-	1	40 (2)	2(2)	1
2	14 (2)	2	5	1	1	1(1)	24 (3)	-	5
3	31(14)	6 (5)	6(3)	-	-	1	44(22)	-	2
4	26 (4)	2	1(1)	1	-	5	35 (5)	-	-
5	65 (6)	6 (1)	2(1)	7(2)	7	-	87(10)	4(2)	2
6	29 (7)	2 (1)	2	-	-	1	34 (8)	4(4)	3
7	39(18)	4 (2)	1	4(2)	-	-	48(22)	1(1)	9
8	52 (3)	11 (5)	6	5	1	2	77 (8)	8(6)	5
9	43 (1)	7 (5)	6(1)	2(1)	-	1	59 (8)	3(2)	5
10	52 (7)	9 (1)	2	1	1	6	71 (8)	5(1)	4
11	29(10)	3 (2)	2	1	-	14(3)	49(15)	1(1)	3
12	49 (4)	2 (1)	6	1(1)	-	-	58 (6)	1(1)	2
13	23 (1)	2 (2)	2	1	4	2	34 (3)	3(2)	4
14	23 (3)	6 (6)	5	1(1)	-	3	38(10)	3(3)	9
15	2	23 (3)	2	4	-	-	3 (3)	1(1)	15
16	1	33 (7)	4(2)	3(1)	3(2)	-	44(12)	6(5)	2
17	27 (4)	5 (1)	3(1)	-	-	3(2)	38 (8)	1	6
18	36 (7)	5 (4)	2(1)	4(2)	-	-	47(14)	4(3)	5
19	10 (1)	3 (2)	7	-	-	1	21 (3)	2(2)	7
20	38 (7)	2 (2)	5(3)	2(2)	-	3	50(14)	3(3)	5
21	61 (6)	11 (6)	-	1(1)	2	4	79(13)	6(6)	1
22	15 (1)	1 (1)	4(3)	3(2)	-	5(4)	28(11)	4(4)	-
23	36 (3)	1 (1)	1	3(1)	-	6(1)	47 (6)	4(1)	4
TOTAL	728	146	84	47	19	59	1083	66	99
TOTAL(*)	(109)	(58)	(16)	(18)	(2)	(11)	(214)	(50)	

	Distance (type no 1)				
SUJET	0	1	2	3+	?
1	3	8	9	7	1
2	8	1	3	2	2
3	10	14	5	2	6
4	6	15	3	2	-
5	14	27	6	18	3
6	17	8	-	4	2
7	11	17	2	9	1
8	4	15	3	30	11
9	2	10	6	25	6
10	7	18	10	17	5
11	6	10	4	9	4
12	1	14	10	24	7
13	10	7	-	6	7
14	6	6	6	5	3
15	12	9	1	1	2
16	4	17	3	9	5
17	8	10	4	5	1
18	10	14	6	6	5
19	4	2	1	3	-
20	4	11	3	20	9
21	3	20	4	34	10
22	7	7	-	1	7
23	9	9	-	18	7
TOTAL	166	269	89	257	104

(*) = formes pronominales erronées (incluses dans le total).

1. Quantité des formes pronominales

Les locuteurs natifs ont un taux nettement plus élevé d'emploi des formes pronominales (la moyenne par sujet : 70 contre 44). Cependant, pour pouvoir en interpréter correctement la signification, il faut prendre en considération le nombre total de termes effectivement susceptibles d'être remplacés (et qui sont remplacés ou non). De façon réaliste, ceci n'est possible que par l'entremise du taux des répétitions. Ici, de nouveau, une différence existe entre les taux des deux groupes : elle est cependant peu importante. En réalité, le taux est très bas pour les deux groupes (3,4 par personne pour les francophones, 4,3 par personne pour les anglophones). Une précision s'impose. Les textes produits par les locuteurs natifs sont, d'habitude (dans nos corpus), relativement

concentrés du point de vue de la quantité des renseignements contenus et l'occasion de remplacer un terme répétitif par une forme pronominale survient alors très souvent; par contre, l'information contenue dans les corpus d'apprenants du français langue seconde est, la plupart du temps, diluée par des propos avortés, par des autocorrections, des reprises purement mécaniques ou par des changements de structures syntaxiques en cours de route, ou encore par des hésitations, etc., ce qui réduit fortement le nombre de termes contenus dans une longueur de texte équivalente. Il convient d'ajouter l'existence d'un deuxième facteur lors de l'analyse du taux assez faible des répétitions dans les deux corpus : dans l'étude, nous avons considéré comme répétitions seulement les reprises intégrales des termes ayant déjà apparus dans les conditions suivantes :

1) La première occurrence du terme se trouvait dans le propos précédant immédiatement celui contenant la répétition ou dans le 2e ou 3e propos (toutes les occurrences suivantes étant traitées comme des nouveaux termes);

2) Une deuxième occurrence du même terme n'a pas été considérée comme une répétition partout où le contexte permettait la possibilité de son interprétation comme un effet de style ou une mise en vedette;

3) De même, on a exclu les reprises du même terme dans les propos contenant déjà une forme pronominale différente si le risque d'ambiguïté entrait en jeu.

Ainsi, seulement les reprises nettement redondantes ont été relevées : le bénéfice du doute a diminué sensiblement le taux de répétitions chez les locuteurs du français langue seconde. Un autre facteur semble confirmer cette supposition : le nombre des répétitions ne comprend pas les reprises du même terme accompagné d'une forme pronominale signalant la volonté du locuteur d'éviter l'ambiguïté (Le rond lui voulait s'en aller...). Les chiffres individuels ne sont pas très élevés en soi, mais il y a une très grande différence entre le taux chez les locuteurs natifs et chez les

locuteurs du français langue seconde : 7,5 par personne et 2,6 par personne respectivement, c'est-à-dire que les locuteurs natifs ont recours à la reprise emphatique presque trois fois plus souvent que les locuteurs anglophones. Le taux de reprises emphatiques (E) comparé avec le taux de répétitions identiques (R) fait également ressortir une différence entre les deux groupes : sur l'ensemble de toutes les répétitions, la répétition avec ajout du pronom emphatique représente chez le locuteur natif 69,2%, mais seulement 37,3% chez le locuteur du français langue seconde. Tout ceci semble corroborer l'hypothèse des répétitions planifiées et fonctionnelles chez les locuteurs natifs, contrairement aux répétitions souvent à caractère aléatoire chez les apprenants du français (tributaire possiblement du manque d'expérience du traitement du système pronominal).

2. Types de formes pronominales : emploi correct et emploi erroné

Les données recueillies sont beaucoup plus frappantes en ce qui concerne les fonctions syntaxiques, tant des termes remplacés par des formes pronominales que des formes pronominales elles-mêmes.

TABLEAU 3 QUANTITÉ DES FORMES PRONOMINALES

Catégorie	Balablok ANG		Balablok FRA	
1	728	71,1%	576	74,7%
2	146	14,3%	47	6,1%
3	84	8,2%	64	8,3%
4	47	4,6%	31	4,0%
5	19	1,9%	53	6,9%
TOTAL	1024	100%	771	100%

Dans les deux groupes c'est de loin la catégorie 1 (terme sujet remplacé par une forme pronominale sujet) qui l'emporte sur les autres catégories : 71,1% (N=728) du total des formes pronominales (N=1024) chez les anglophones, 74,7% (N=576) du total des formes pronominales (N=771) chez les francophones. Pour les autres catégories, les taux sont assez faibles dans les deux groupes. Cependant, pour ce qui est des catégories 2 à 5 respectivement, il y a une nette différence entre les deux groupes. Chez

les anglophones, la catégorie 2 est en 2e place avec 14,3% (N=146). Suivent les catégories 3 (N=84), 4 (N=47), 5 (N=19) dont les taux baissent respectivement environ de moitié par rapport à la catégorie précédente (8,2%, 4,6%, 1,9%). Chez les francophones, la 2e catégorie la plus représentée est la catégorie 3, (N=64), suivie des catégories 5 (N=53) et 2 (N=47) et finalement de la catégorie 4 (N=31) avec 8,3%, 6,9%, 6,1% et 4% respectivement.

Reste à considérer le taux d'emploi des formes pronominales par rapport aux taux de réussite, c.-à-d. l'emploi correct vis-à-vis de l'emploi erroné des pronoms. Chez les locuteurs natifs, aucune erreur n'a été relevée, c.-à-d. le taux de réussite se situe à 100%. Pour le groupe anglophone, le taux d'erreurs dans les catégories 2, 3 et 4 est considérable : 30% (N=92) pour l'ensemble des catégories, avec le taux le moins élevé pour la catégorie 3 (coréférent objet - forme pronominale sujet; 19%, N=16) et le taux le plus élevé pour la catégorie 2 (coréférent sujet - forme pronominale objet; 39,7%, N=58) suivie de près par la catégorie 4 (coréférent objet - forme pronominale objet; 38,3%, N=18). Le taux d'erreurs est cependant significatif même pour la catégorie 1 (coréférent sujet - forme pronominale sujet) : 15% (N=109).

TABLEAU 4 TAUX D'ERREURS

Catégorie	Balablok ANG		Balablok FRA	
1	15,0%*	109	0%*	0
2	39,7%	58	0%	0
3	19,0%	16	0%	0
4	38,3%	18	0%	0
5	10,5%	2	0%	0
TOTAL	19,8%**	203	0%**	0

*100% = total individuel d'occurrences pour chaque catégorie
**100% = total d'occurrences pour l'ensemble des catégories 1, 2, 3, 4, 5
ANG : 100% = 1024
FRA : 100% = 771

Les taux d'erreurs indiquent le peu d'assurance des locuteurs du français langue seconde vis-à-vis de l'emploi des formes de pronoms personnels à la fonction syntaxique d'objet, soit direct, soit indirect. La plupart de ces erreurs réduisent la clarté du

message : cependant, dans la version transcrite, il est encore possible de comprendre le sens par le contexte. S'ajoutent au taux général d'erreurs les cas où il est tout à fait impossible de déterminer à l'aide du contexte quel terme est le coréférent de la forme pronominale (voir le Tableau 2, la colonne ?, N = 104, soit 4,5 par locuteur). De plus, pour les propos contenant deux formes pronominales ayant deux coréférents différents (même Tableau, colonne 2FP(*)), le taux d'erreurs s'élève à 76% (N = 50).

3. Distance entre la forme pronominale et son coréférent

Le facteur de la distance entre les deux coréférents atteste lui aussi des différences significatives entre les deux groupes de locuteurs. Pour les locuteurs anglophones, le taux des formes pronominales (pour la catégorie 1) contenues dans le même propos que leurs coréférents respectifs s'élève à 21% (N = 166) comparé à 12% (N = 70) pour les locuteurs natifs. Pour les coréférents apparaissant dans deux propos contigus, les taux sont comparables pour les deux groupes : 34% (N = 269) et 27% (N = 157) respectivement. Pour les coréférents séparés par deux séparateurs d'unités de base (les propos entreposés peuvent contenir ou non une forme pronominale ayant le même coréférent), le taux baisse pour les deux groupes mais plus radicalement chez les locuteurs anglophones (11% - N = 89 - contre 13% - N = 75 - chez les francophones). La plus grande différence se situe, cependant, au niveau des coréférents les plus éloignés les uns des autres (3e propos ou plus). Tandis que chez les anglophones le taux se situe à 33% (N = 257), il s'élève à 48% (N = 274) chez les francophones.

TABLEAU 5 DISTANCE ENTRE LA FORME PRONOMINALE ET SON CORÉFÉRENT (CATÉGORIE 1)

Distance	Balablok ANG		Balablok FRA	
0	166	21,3%*	70	12,2%
1	269	34,4%	157	27,3%
2	89	11,4%	75	13,0%
3	257	32,9%	274	47,6%
TOTAL	781	100%	576	100%

*100% = total des catégories 0, 1, 2, 3

Les modèles d'emploi des coréférents sont ainsi assez différents pour les deux groupes de locuteurs : tandis que les francophones tendent à l'économie d'expression (par remplacement des termes par des formes pronominales dans des propos consécutifs à travers des étendues considérables de texte), les locuteurs anglophones semblent employer la plupart des formes pronominales à des distances beaucoup moins importantes.

Conclusion

Pour résumer, nous croyons avoir démontré l'existence de plusieurs différences significatives dans l'emploi des formes pronominales chez les locuteurs natifs et chez les apprenants du français langue seconde. Premièrement, il nous semble évident que l'emploi correct des formes pronominales témoigne d'une certaine excellence du locuteur : le taux d'erreurs élevé caractérise souvent des locuteurs faibles et diminue considérablement la clarté d'expression. Les locuteurs du français langue seconde semblent employer les formes pronominales relativement fréquemment (quoique moins que les locuteurs natifs), mais en faisant souvent des erreurs. Les locuteurs natifs emploient les formes pronominales pour des raisons d'économie, souvent à des longues distances, sans compromettre néanmoins la clarté d'expression, et témoignent ainsi d'une production linguistique contrôlée et dirigée. Par contre, les apprenants du français langue seconde emploient les formes pronominales surtout à courte distance. Aussi, leur emploi des formes pronominales ne semble pas obéir à des règles explicites mais relève plutôt du degré d'assurance linguistique du locuteur.

Nous imputons ces différences principales à la préférence marquée chez les apprenants pour la manipulation de termes plus explicites (les formes pronominales étant plus abstraites) d'une part et aux lacunes dans leurs connaissances grammaticales de la langue seconde d'autre part.

Références

Bosch, P. (1983). *Agreement and anaphora*. Londres : Academic Press.

Cornish, F. (1986). *Anaphoric relations in English and French : a discourse perspective*. Beckenham : Croom Helm Ltd.

Dahl, O. (dir.) (1974). *Topic and comment, contextual boundness, and focus*. Hambourg : Buske.

De Beaugrande, R. (1980a). *Text, discourse, and process*. Norwood, N.J. : Ablex; Londres : Longman.

De Beaugrande, R. et W. Dressler (1981). *Introduction to text linguistics*. Londres : Longman Group Ltd.

Fauconnier, G. (1974). *La coréférence : syntaxe ou sémantique?* Paris : Seuil.

Halliday, M. et R. Hasan (1976). *Cohesion in English*. Londres : Longman.

Hobbs, J. (1979). "Coherence and co-reference". dans *Cog.Sci.*, 3, 67-90.

Lakoff, G. (1968). *Pronouns and reference*. Bloomington : Indiana University Linguistics Club.

Lust, B. (éd.) (1986). *Studies in the acquisition of anaphora*, vol. 1. Dordrecht : D. Reidel Publishing Company.

Schecker, M. et P. Wunderli (dirs.) (1975). *Textgrammatik : Beiträge zum Problem der Textualität*. Tübingen : Niemeyer.

Smyth, R.H. (1986). *Cognitive aspects of anaphora judgment and resolution*. Bloomington : Indiana University Linguistics Club.

Tannen, D. (éd.) (1984). *Coherence in spoken and written discourse*. Norwood, N.J. : Ablex Publishing Corporation.

Quelques aspects de la cohésion textuelle : paraphrase, récurrence, parallélisme, ellipse

Béatrice Majkrak

Les deux ensembles de corpus Balablok ont un dénominateur commun : toute la production orale de chacun des locuteurs évolue autour d'un même sujet. Il s'agit de raconter le contenu d'un court métrage d'animation. Les détails du déroulement de l'action racontée étant facilement vérifiables, il est possible de porter un jugement sur le degré de précision ou d'ambiguïté de chaque énoncé, ainsi que sur le degré d'importance de chaque détail mentionné.

L'observation du comportement verbal du groupe contrôle (locuteurs natifs) permet de vérifier dans quelle mesure l'emploi et l'agencement des renseignements à l'intérieur d'une même étendue de texte est prévisible, c.-à-d. quels sont les procédés collectifs visant à rendre le message compréhensible.

Pour la compréhension d'une courte étendue de texte parlé, il faut, outre la connaissance du système linguistique impliqué, pouvoir s'appuyer sur des renseignements obtenus avant ou après (à l'intérieur du même texte cependant). Le processus d'encodage des suites de sens et des réseaux de relations entre différents éléments constituant le message accuse sans doute des traits collectifs qu'on peut imputer aux conventions non explicites. Autrement dit, les modèles usuels du traitement linguistique des

renseignements par les locuteurs sont en quelque sorte le reflet de l'efficacité de ces premiers du point de vue de l'interlocuteur. De telles stratégies (indépendantes des servitudes grammaticales) sont probablement communes à toutes les langues et ne devraient pas constituer un problème pour les locuteurs des langues secondes, surtout quand il s'agit de langues à contextes culturels très semblables, comme c'est le cas de l'anglais et du français.

Cependant, un aspect important différencie les deux groupes de locuteurs étudiés. Pour un locuteur natif, l'attention portée à l'organisation des renseignements peut être la tâche principale lors du récit, la structuration purement syntaxique étant un exercice pour ainsi dire automatique et qui ne pose pas de problèmes. Pour un apprenant de langue seconde, il s'agit plus probablement d'une tâche additionnelle, d'importance secondaire, puisqu'il semble porter toute son attention tout d'abord sur la grammaire et le vocabulaire, d'autant plus s'il n'a pas une connaissance approfondie de la langue. Il est d'ailleurs à supposer qu'un locuteur parlant une langue seconde pensera à organiser la structure de surface de ses propos à court terme, c.-à-d. aux problèmes relevant directement de sa propre compétence, avant d'aborder la question de la capacité de son interlocuteur à suivre l'enchaînement logique des renseignements contenus dans ces propos.

Le degré de compréhensibilité d'un texte peut varier en fonction de l'anticipation, de la part du locuteur, de la capacité de son interlocuteur à comprendre non seulement chaque mot ou séquence de mots, mais aussi à saisir les liens entre les différentes parties du texte. Un locuteur expérimenté saura ainsi par exemple évaluer les limites de la mémoire à court terme de son interlocuteur. En conséquence, il va tantôt répéter certaines parties, soit textuellement, soit par une paraphrase, tantôt accentuer un mot ou une séquence ou bien encore recourir à l'emploi des structures syntaxiques facilitant à son avis la compréhension. Dans tous ces cas, il s'agira de l'emploi arbitraire de certains procédés basés sur le choix de structures syntaxiques, sur le choix de modèles lexicaux et sur leur distribution à travers des étendues du texte plus ou moins importantes, c.-à-d. de procédés cohésifs.

Deux résultats opposés peuvent être visés et obtenus par l'emploi de procédés cohésifs : l'économie d'expression d'une part

et le caractère explicite et univoque de l'énoncé d'autre part. Nous avons déjà abordé la question des formes pronominales, un moyen largement employé pour raccourcir la structure de surface par l'emploi, le plus souvent, de pronoms remplaçant des expressions plus longues et plus explicites. Si le remplacement d'un mot par une forme pronominale teste d'une certaine façon la mémoire de l'interlocuteur, plusieurs procédés sont employés au contraire dans le but d'aider ce dernier dans sa tâche de retenir des faits déjà mentionnés et de lui en faciliter le décodage. Il va sans dire que de tels procédés diminuent l'économie d'expression, étant souvent basés sur la répétition de certains éléments déjà employés. Nous avons étudié plusieurs de ces procédés : la récurrence, le parallélisme et la paraphrase. Pour compléter l'étude de l'économie d'expression, nous avons également relevé les occurrences d'ellipse.

1. **Paraphrase** : occurrence du même sens dans une expression différente

   - je me souviens plus de ça - J'AI DÉJÀ OUBLIÉ CELA
   - ils se saluent avec leurs chapeaux - ILS ENLÈVENT LEUR CHAPEAU - BONJOUR BONJOUR
   - ils se parlent - c'est en dessin - TU VOIS LE DESSIN DE CE QU'ILS VEULENT FAIRE

2. **Récurrence** : la répétition des mêmes mots ou des mêmes expressions

   - c'est plus rond - ... - C'EST PLUS ROND
   - ils sont toutes devenus des hexagones - il y avait juste un petit carré jaune - il essayait de trouver un rond ... - il en avait plus - ILS ÉTAIENT TOUTES DEVENUS DES HEXAGONES
   - au début ça avait l'air plus d'un jeu - ç'avait l'air innocent ... - ils se mettent à jouer à la balle avec - ... Ç'AVAIT L'AIR INNOCENT

3. **Parallélisme** (récurrence syntaxique) : une structure syntaxique répétée mais contenant de nouveaux éléments lexicaux

   - ils ne disaient pas bonjour mais ILS DISAIENT QUELQUE CHOSE
   - la sphère est repartie - LES CARRÉS SONT REPARTIS
   - lui c'est un carré - LES AUTRES C'EST DES SPHÈRES

4. **Ellipse** : une structure répétée partiellement tout en maintenant son contenu, c.-à-d. où on omet une (ou plusieurs) expression(s) récupérable(s) par le contexte

   - ils les bousculent - (ILS LES) frappent - (ILS) leur donnent des coups de pied
   - c'est comme entre les Français et les Anglais - les Turcs et les Arméniens - les Juifs et les Arabes

Méthodologie

Les deux corpus Babablok se prêtent particulièrement bien à l'étude du traitement linguistique des renseignements transmis, chacun des locuteurs ayant pour tâche de décrire le contenu d'un même court métrage d'animation. Il s'agit tout d'abord d'expliquer l'action du film (rencontre de deux formes géométriques différentes, bataille entre les deux groupes, transformation des deux formes en une autre forme géométrique menant à la fin des hostilités) et de répondre par la suite à des questions portant sur des détails précis de son déroulement. Pour mettre en évidence le message véhiculé par le film (la coexistence pacifique entre les races différentes est possible), il a fallu énumérer plusieurs étapes du conflit et décrire plusieurs scènes parfois assez compliquées. Tout au long du récit, il était ainsi important de manipuler plusieurs données significatives et de faire référence à plusieurs personnages, événements et objets ayant un rapport les uns avec les autres à des degrés différents.

Nous avons procédé en deux étapes :

Dans chaque corpus, des «thèmes» ont été isolés. Chaque thème contient un seul renseignement précis, plus ou moins développé à l'aide de détails (description d'un type de personnage, rencontre des carrés et des balles, technique particulière de combat, etc.). Ces thèmes apparaissent tantôt spontanément, surtout au début de l'entrevue, tantôt en guise de réponse à une question de la part de l'interviewer.

Chaque corpus a été étudié en vue de relever les cas d'occurrences d'ellipses, de récurrences, de parallélismes et de paraphrases. Nous avons pris le soin de ne relever que les cas où l'occurrence apparaît à l'intérieur d'un même thème et, moins souvent, les cas d'occurrences dans deux thèmes contigus et apparentés par le sens. Nous n'avons pas retenu les cas de structures répétitives ou parallèles avortées ou incomplètes, lesquelles, d'ailleurs, auraient pu être interprétées différemment, une fois complétées.

Compilation et analyse des données

Les relevés d'occurrences de tous les phénomènes étudiés apparaissent dans les tableaux suivants.

TABLEAU 1 BALABLOK ANGLOPHONES

SUJET	N° Thèmes (TP)	Pp	Re	Pl	El
1	33(31)	6	3	2	2
2	36(20)	6	3	-	2
3	58(40)	6	7	2	-
4	45(37)	3	1	2	2
5	33(28)	6	1	4	1
6	33(30)	5	2	-	1
7	27(25)	1	3	6	4
8	24(17)	4	5	8	-
9	37(36)	2	2	4	5
10	25(21)	4	1	8	1
11	33(27)	2	1	1	1
12	23(17)	5	4	7	1
13	21(20)	-	2	5	2
14	25(21)	1	5	2	2
15	28(23)	1	1	3	1
16	15(14)	-	1	6	-
17	27(26)	-	-	11	-
18	31(28)	1	2	2	1
19	33(33)	2	1	1	1
20	18(16)	5	5	4	1
21	15(10)	8	6	3	2
22	23(21)	1	1	2	2
23	30(23)	1	3	1	-
TOTAL	673(564)	70	60	84	32
MOYENNE	29,31(24,5)	3,0	2,6	3,7	1,4

Pp = paraphrase
Re = récurrence
Pl = parallélisme
El = ellipse
TP = thème provoqué (chiffre partiel)

TABLEAU 2 BALABLOK FRANCOPHONES

SUJET	N° Thèmes (TP)	Pp	Re	Pl	El
1	34(18)	13	8	13	3
2	38(34)	16	11	2	3
3	35(28)	11	6	2	-
4	39 (6)	16	9	15	1
5	47(14)	21	6	6	-
6	48(29)	12	6	4	2
7	44(26)	21	8	3	2
8	34(16)	13	12	9	2
9	31(23)	22	11	6	3
10	50 (5)	8	2	5	2
11	30 (4)	26	13	3	6
TOTAL	430(203)	173	92	68	24
MOYENNE	39,1(18,5)	15,7	8,4	6,2	2,2

Pp = paraphrase
Re = récurrence
Pl = parallélisme
El = ellipse
TP = thème provoqué (chiffre partiel)

Les données chiffrées, d'une part, et les observations faites tout au long des lectures répétées des corpus, d'autre part, nous permettent de tirer quelques conclusions :

Premièrement, pour une même longueur absolue de texte (100 lignes de texte suivi, à l'exclusion des interventions de l'interviewer), les locuteurs natifs ont abordé un nombre nettement plus élevé de thèmes (39 thèmes par locuteur contre 29 thèmes par locuteur chez les apprenants du français langue seconde). De même, les locuteurs natifs ont un taux plus élevé pour des thèmes abordés spontanément (opposés aux thèmes enclenchés par une question) : 53% du total des thèmes, c.-à-d. 21 thèmes sur 39 pour les francophones comparativement à 16% du total des thèmes chez les anglophones, soit 5 thèmes sur 29. Ceci se traduit par une quantité accrue de renseignements fournis par le locuteur de façon spontanée. Les différents thèmes sont reliés, soit parce qu'ils se suivent dans l'histoire racontée, soit par une logique interne. Les apprenants du français langue seconde répondent la plupart du temps à des questions : les thèmes

apparaissent nettement détachés les uns des autres et les liens logiques sont moins évidents.

TABLEAU 3 THÈMES ABORDÉS

Thèmes	FRA*	ANG*
provoqués	18 47%	24 84%
spontanés	21 53%	5 16%
TOTAL	39 100%	29 100%

* = taux calculés par locuteur

Deuxièmement, le taux d'occurrences de l'ellipse est numériquement peu significatif dans les deux ensembles de corpus mais, de nouveau, il est plus élevé chez les locuteurs natifs (2,2 par locuteur pour les francophones et 1,4 par locuteur pour les anglophones). Les occurrences relevées se limitent la plupart du temps aux cas de l'ellipse du verbe (ex. : ils ont pensé à s'en servir - à en abuser) ou - moins souvent et surtout dans les corpus anglophones - de l'ellipse du sujet (ex. : ils se mettaient ensemble - parlaient - discutaient).

Ensuite, les taux comparés, pour les deux groupes de locuteurs, de paraphrases, de récurrences et de parallélismes sont par contre significatifs : dans les trois cas, ce sont les locuteurs natifs qui accusent un taux nettement plus élevé. La différence la plus grande se situe au niveau de la paraphrase : 15,7 occurrences par locuteur chez les francophones contre trois occurrences seulement chez les anglophones, c.-à-d. qu'un locuteur natif emploie cinq fois plus la paraphrase qu'un apprenant du français langue seconde. Pour la récurrence et le parallélisme, la différence est encore grande : 8,4 et 6,2 respectivement pour les locuteurs natifs comparativement à 2,6 et 3,7 respectivement pour les anglophones, c.-à-d. environ trois et deux fois plus d'occurrences respectivement chez les premiers. En tenant compte de la proximité des catégories de paraphrase et de récurrence (répétition d'une même idée), nous pouvons dire que pour 1,6 thèmes il y a, chez le locuteur natif, une répétition, c.-à-d. une mise en évidence ou un rappel des points saillants. Pour les anglophones, ceci n'est vrai que dans la mesure d'un rappel pour 5,2 thèmes. Il est à

supposer que l'emploi de la paraphrase implique une connaissance approfondie des synonymes et des nuances lexicales en général. Ceci pourrait expliquer en partie le taux très bas de paraphrases chez les apprenants du français langue seconde. Comme cependant les taux des répétitions textuelles et syntaxiques (parallélismes) sont également très bas chez les anglophones, il est logique de supposer que ces procédés cohésifs sont mal maîtrisés par les apprenants du français langue seconde.

Enfin, des différences entre les deux groupes ressortent aussi d'une comparaison effectuée du point de vue de l'importance de chacun des quatre moyens cohésifs relativement à l'ensemble de leur emploi. Pour le groupe francophone, le moyen le plus utilisé est la paraphrase (48,3%). Suivent la récurrence (25,8%) et le parallélisme (19,1%). L'ellipse ne représente que 6,8% du total. Chez les locuteurs anglophones, le parallélisme occupe la première place avec 34%, la paraphrase et la récurrence suivent d'assez près (28,3%, 24,5%) et l'ellipse occupe également la quatrième position avec 13,2%.

Le seul moyen cohésif qui est ainsi comparable de ce point de vue dans les deux groupes est la récurrence. Le parallélisme et l'ellipse sont presque deux fois plus importants pour les anglophones que pour les francophones. Par contre, la paraphrase est de loin le moyen cohésif le plus employé par les locuteurs natifs, tandis que pour les anglophones, elle est utilisée presque aussi peu que la récurrence et le parallélisme.

TABLEAU 4 OCCURRENCES ET IMPORTANCE RELATIVE DES QUATRE MOYENS COHÉSIFS ÉTUDIÉS

Moyen cohésif	FRA*		ANG*	
Ellipse	2,2	6,8%	1,4	13,2%
Paraphrase	15,7	48,3%	3	28,3%
Récurrence	8,4	25,8%	2,6	24,5%
Parallélisme	6,2	19,1%	3,7	34%
TOTAL	32,5	100%	10,7	100%

* = taux calculés par locuteur

Ensemble, les données montrent assez clairement que les locuteurs natifs tendent vers l'explicitation du message par modification du matériel linguistique, tandis que les apprenants du français langue seconde, ayant probablement un répertoire moins varié d'unités lexicales et de structures syntaxiques, ont plutôt recours à la répétition qui leur demande moins d'effort.

Conclusion

La comparaison de plusieurs procédés contribuant à la cohésion du texte, dans les deux ensembles de corpus, a fait ressortir des différences significatives au niveau de l'organisation des renseignements. Les taux d'occurrences sont plus élevés chez les francophones dans toutes les moyennes et dans la majorité écrasante des scores individuels. Les taux, quoiqu'accusant une certaine variation à l'intérieur du même groupe linguistique, permettent de supposer l'existence de certaines constantes : l'emploi, relativement systématique et de toute évidence intentionnel, des procédés cohésifs chez les locuteurs natifs, d'une part, et l'utilisation fort peu significative de ces mêmes procédés à travers les corpus des apprenants du français langue seconde, d'autre part.

La répétition des mêmes éléments implique nécessairement une diminution des renseignements contenus dans une étendue de texte donnée et entraîne une certaine monotonie d'expression. Les taux plus élevés de paraphrases et de récurrences, d'une part, et de parallélismes, d'autre part, devraient ainsi aboutir, pour une longueur égale de texte, à un contenu plus dilué et à un degré de redondance plus élevé chez les francophones, comparativement aux apprenants du français langue seconde. Or, il n'en est rien. Le premier indice en est d'ailleurs le nombre de thèmes abordés : en général, un locuteur natif inclut dans son récit plus de renseignements partiels composant le message global. Quoique la longueur de chaque thème soit moins importante pour les francophones que pour les anglophones, les renseignements contenus y sont plus détaillés, tout en comportant souvent une répétition qui rafraîchit la mémoire de l'interlocuteur ou fait ressortir les points importants. Ceci est possible grâce au fait qu'il n'y a pas de perte d'espace, due aux répétitions non fonctionnelles de mots grammaticaux (articles, verbes auxiliaires, etc.), aux hésitations, aux autocorrections, aux structures avortées en cours de route et aux

formulations maladroites, trop longues et au sens vague, faute de connaissance de termes précis. Tous ces phénomènes sont par contre très fréquents chez les apprenants du français langue seconde. Ainsi, chez ces derniers, l'information transmise contient peu de détails, les faits sont constatés une seule fois et de façon peu explicite. Il est difficile d'imputer au texte l'importance de renseignements partiels : s'agit-il d'un détail sans conséquence ou d'un point central de l'événement? Les locuteurs anglophones ne semblent pas tenir compte de cet aspect important de la communication : dans leur souci de respecter tout d'abord les règles grammaticales de base, ils oublient totalement de structurer logiquement leur récit. Très souvent, il est ainsi possible de saisir le contenu d'un passage seulement en relisant la transcription : à l'oral, le renseignement est souvent irrémédiablement noyé dans une séquence monotone dont la ponctuation n'est pas dictée par la logique mais plutôt par le degré de difficulté des structures grammaticales employées.

Nous croyons ainsi justifié de prétendre que certains procédés augmentant la cohésion textuelle, telle la récurrence, la paraphrase et le parallélisme semblent faire partie des stratégies employées par les locuteurs natifs dans le but de rendre leur récit plus facile à comprendre. La répétition littérale ou la paraphrase mettent ainsi en vedette certains points jugés importants, d'une part, et le parallélisme de structures syntaxiques facilite la compréhension du sens en le présentant dans des structures formelles identiques, d'autre part. Quoique ne relevant pas des servitudes grammaticales, ces procédés cohésifs semblent faire partie d'une convention non explicite de la communication entre les membres d'un même groupe linguistique. Leur absence chez les apprenants du français langue seconde contribue à la désorientation de l'interlocuteur face aux renseignements transmis.

Références

Dahl, O. (dir.) (1974). *Topic and comment, contextual boundness and focus*. Hambourg, Buske.

De Beaugrande, R. (1980). *Text, discourse and process*. Norwood, N.J., Ablex; Londres, Longman.

De Beaugrande, R. et W. Dressler (1981). *Introduction to text linguistics*. Londres, Longman.

Fauconnier, G. (1974). *La coréférence : syntaxe ou sémantique?* Paris, Seuil.

Halliday, M. et R. Hasan (1976). *Cohesion in English*. Londres, Longman.

Hasan, R. (1968). *Grammatical cohesion in spoken and written English*. Londres, Longman.

Lundquist, L. (1980). *La cohérence textuelle : syntaxe, sémantique, pragmatique*. Copenhague, Nyt Nordish Forlag.

Schecker, M. et P. Wunderli (dirs.) (1975). *Textgrammatik : Beiträge zum Problem der Textualität*. Tübingen : Niemeyer.

Tannen, D. (éd.) (1984). *Coherence in spoken and written discourse*. Norwood, N.J. : Ablex Publishing Corporation.

Ungeheuer, G. (1969). *Paraphrase und syntaktische Tiefenstruktur*. Folia Linguistica 3, 178-227.

Étude corrélative de la variation grammaticale et discursive

Kathleen Connors

Introduction

La présente étude traite de la comparaison du français de jeunes adultes anglophones et d'immigrants adultes au Québec : dans le contexte de notre recherche, les lusophones et les allophones dont nous parlons dans l'*Introduction*. Nous savons que les jeunes anglophones québécois apprennent le français principalement à l'école, tandis que les immigrants l'apprennent davantage au travail et «dans la rue», bien que la plupart des locuteurs dont nous étudions le français ici aient suivi au moins un cours de français ou même, dans certains cas, aient été scolarisés partiellement en français.

L'apprentissage et/ou l'acquisition du français, surtout chez les adultes et tout particulièrement chez les immigrants, n'ont pas vraiment été étudiés en détail (Trévise et Noyau 1984; Véronique 1984). Dans le domaine plus général de la recherche sur l'acquisition des langues secondes, le nombre même des études sur la nature du français L2 n'est pas comparable au nombre consacré à l'anglais langue seconde. Nous voudrions commencer à combler cette lacune, non seulement au nom de la description des langues pour elle-même, mais aussi parce que certains problèmes classiques de la structure du français, décrits par la grammaire traditionnelle et moderne, sont aussi des problèmes concrets pour les apprenants adultes du français.

Le cas le plus évident, peut-être, du point de vue des particularités typologiques du français par opposition à l'anglais, serait le système des pronoms clitiques compléments, dont la morphosyntaxe n'est maîtrisée que très rarement par les anglophones, même si leur connaissance passive du français est d'un bon niveau par ailleurs. Nous avons traité de ce problème dans un autre article (Connors et Nuckle 1986), et fixons notre attention cette fois-ci sur un autre ensemble de problèmes structuraux.

À un niveau plus général, nous nous intéressons aux aspects des L2 qui «résistent» à l'acquisition par les adultes. Nous postulons que les effets de cette résistance s'exercent de façon différentielle, de sorte que certains groupes de locuteurs adultes apprennent ces structures ou ces traits problématiques mieux que d'autres groupes. Les raisons du caractère différentiel de ces effets pourraient être génétiques (au sens linguistique) ou typologiques. Ou, au contraire, elles pourraient s'apparenter au type d'acquisition ou d'apprentissage que le groupe a connu. Dans le cas de nos informateurs, ces deux genres de causes sont malheureusement confondus.

La phase plus spécifique de notre étude que nous voulons présenter ici est une tentative d'éclaircissement des rapports entre types de compétence linguistique chez les locuteurs d'une L2, tels que perçus à travers des genres différents d'évaluation de cette compétence. Le premier type d'évaluation est structural, puisqu'il s'agit d'évaluer des choix de formes déterminées par la morphologie, la syntaxe et la sémantique. Cette évaluation est classique également en ce sens qu'elle se base sur une analyse des erreurs, en la quantifiant. Il s'agit du degré auquel chaque informateur aura exploité le système des temps, des aspects et des modes en français par rapport à la façon de faire d'un francophone en ce domaine.

La deuxième évaluation en jeu ici est extra-structurale, au sens étroit. Elle tient compte d'éléments en dehors des propositions (en prenant la proposition comme unité fondamentale de l'analyse morphosyntaxique). C'est une tentative de trouver un indice superficiel du degré de ce qu'on pourrait appeler la compétence «discursive», tenue naïvement ici pour la capacité d'entretenir et de poursuivre le discours en liant, et même, en interrompant les

propositions par une variété d'expressions. Il est clair qu'il faudra défendre cette optique ci-dessous.

Le troisième type d'évaluation est totalement «extra-structural», en ce sens que son exécution ne dépend pas de notre capacité d'analyser les phrases, les propositions et les syntagmes des corpus de nos informateurs. Il s'agit d'évaluations globales, par des francophones, du discours oral des apprenants. Nous nous intéressons à de telles appréciations comme moyen de vérification : avons-nous tort ou raison d'opposer les deux types d'évaluation linguistique mentionnés ci-dessus? Nous vérifierons le caractère distinct que nous leur avons implicitement attribué, en les comparant tour à tour avec l'évaluation globale. Ensuite, nous comparerons deux sortes d'évaluations globales, pour ce qui est de leurs correspondances aux deux évaluations linguistiques. Ci-après, nous distinguons sommairement les types d'évaluations globales comme [± «naïve»].

La question sous-jacente qui nous pousse à comparer la nature et les résultats d'évaluations, tenues *a priori* pour divergentes, est celle de savoir si la compétence linguistique est unitaire, ou, au contraire, est un ensemble de capacités distinctes qu'il conviendrait d'opposer comme «grammaticales» et «communicatives». On invoque souvent de tels termes pour expliquer les cas où les divers types de mesures linguistiques aboutissent à des évaluations divergentes de la compétence relative des apprenants/acquérants d'une L2. Il en fut ainsi, par exemple, lors de certains de nos travaux antérieurs (Connors 1983). Plus généralement, la littérature sur «l'immersion française», à laquelle nous reviendrons ci-dessous, a démontré que les évaluations globales situent les participants à de tels programmes beaucoup plus près des témoins francophones que ne le font les évaluations analytiques, c'est-à-dire celles qui exigent l'isolement d'unités linguistiques (Connors 1987).

Ici, nous poserons la question de la distinction entre types de capacités linguistiques, telle que reflétée dans la comparaison des évaluations d'apprenants : indépendamment de la distinction entre évaluations globales et analytiques, est-ce que les évaluations axées sur des niveaux distincts d'analyse linguistique diffèrent aussi dans leur degré de correspondance aux évaluations autoch-

tones? L'avantage principal des mesures opposées ci-dessous est vraisemblablement le fait qu'elles soient intuitivement associées à la «grammaire», d'un côté, et au «discours», de l'autre. Nous avons essayé, toutefois, de nous limiter ici à deux ensembles de traits linguistiques qui sont très fréquents, grâce à leur statut obligatoire dans la structure de la proposition, d'une part, et dans celle du discours, d'autre part.

Les composantes de l'étude : les corpus utilisés

Pour la présente étude, nous avons utilisé trois des cinq corpus décrits dans l'*Introduction* : les corpus «lusophones», «anglophones» et «allophones», puisqu'il s'agit principalement de comparer nos apprenants du français L2 entre eux, à l'intérieur de chaque groupe ayant subi une entrevue donnée. Indirectement, nous pourrons comparer les trois groupes d'apprenants entre eux, pour ce qui est des indices linguistiques et des mesures du niveau d'acquisition qui nous intéressent. Pour les besoins de l'étude statistique dont il sera question plus bas, il faut se rappeler (voir l'*Introduction*) que le corpus «allophone» n'a pas été informatisé. Étant donné qu'une des variables que nous voudrons mettre en rapport avec nos indices linguistiques est la longueur du corpus, nous avons compté les mots (orthographiques) du corpus allophone à la main. Même si nous avons essayé d'appliquer exactement les mêmes conventions que celles employées par le dénombrement automatique des mots chez les lusophones et les anglophones, il faut considérer le décompte des mots comme approximatif dans le cas des allophones. Pour cette raison, et à cause des différences importantes entre les groupes d'informateurs et les entrevues qu'ils ont subies (voir l'*Introduction*), nous scinderons l'étude présente en trois phases, correspondant aux «lusophones», aux «anglophones» et aux «allophones», même si certaines comparaisons indirectes entre les trois ensembles de données auront une importance capitale pour la conclusion ci-dessous.

Analyses et évaluations

Nous voudrions indiquer ici, dans un premier temps, les données linguistiques que nous avons isolées afin de formuler des mesures d'évaluation.

A. Les formes verbales

Toutes les occurrences des formes verbales à temps fini dans tous les corpus mentionnés ci-dessus, ainsi que les formes infinitives qui ne font pas partie d'un futur périphrastique et les participes à l'exclusion de ceux contenus dans les temps composés, ont reçu un code (normalement un nombre entre 0 et 9). Dans le cas de nos corpus informatisés, l'extraction des formes encodées, avec leur contexte immédiatement à gauche et à droite, nous permet de dénombrer et d'évaluer la grammaticalité des occurrences des divers sous-types. Les codes différencient les formes uniquement en fonction du temps-mode-aspect, cependant. Ils n'identifient pas les autres traits morphologiques de la forme verbale et ne servent pas à une analyse globale des sources d'erreurs (sur le thème verbal, par exemple). Le code suit la forme verbale directement et est immédiatement suivi d'un astérisque et un code correcteur, si l'analyste francophone juge que la forme est agrammaticale, dans son contexte, pour ce qui est du temps-mode-aspect. Le cas échéant, l'astérisque est suivi d'un second code indiquant le temps-mode-aspect par lequel l'analyste remplacerait la forme erronée. Ainsi, *parle* 1*3 indiquerait que le sujet avait prononcé une forme du verbe *parler* au temps présent de l'indicatif, alors qu'il aurait dû employer un passé composé dans ce contexte.

B. Les éléments extra-propositionnels

Nous expérimentons une approche de l'évaluation superficielle de la performance linguistique au niveau discursif : une évaluation de la variété (en dénombrant les types lexicaux) des connecteurs et des incises. Nous employons ce dernier terme (en anglais aussi) pour désigner les expressions parenthétiques qui sont courtes et figées. Le concept «connecteur» désigne ici les mots ou expressions figées qui lient les phrases et les propositions, par opposition aux subordinateurs, d'un côté, et aux conjonctions de syntagmes, de l'autre. De toute façon, ces dernières constituent un sous-ensemble lexical des connecteurs propositionnels. Ce sont les connecteurs «hors construction» (selon la grammaire traditionnelle) que nous essayons d'isoler, c'est-à-dire en dehors de la structure propositionnelle. C'est dans ce sens-là qu'ils fonctionnent au niveau du discours. La grammaire traditionnelle désigne au moins *et*, *mais*, *ou*, *alors* et *donc* de cette façon, dont les trois premiers

seraient des conjonctions de coordination. De fait, *ou*, dans notre corpus, se limite presque entièrement à la fonction de lier les syntagmes. Par contre, l'influence du français parlé et populaire (y compris le québécois) ajoute *puis* à l'inventaire des connecteurs, ainsi que sa variante non-standard */pi/* (laquelle peut être lexicalement distincte pour certains apprenants), et finalement des entrées composées, telles que *et /pi/*, ainsi que *et puis alors*.

Les incises (telles que définies ci-dessus) sont aussi «hors construction», en ce sens qu'elles interrompent la proposition sans la changer syntaxiquement, ou bien qu'elles en sont détachées à droite. En réalité, la catégorie d'incises recoupe considérablement celle des connecteurs. Il y a cependant des mots ou des expressions qui sont des incises par excellence et pas normalement des connecteurs, tels que *je pense* (lorsque ce n'est pas suivi d'un complémenteur). Il y a également, bien entendu, des lexèmes qui sont grammaticaux seulement dans la fonction de connecteur, et non comme incise, tels que *et*, *mais*, *ou*, *puis*, les conjonctions de coordination.

Un exemple typique du chevauchement des catégories de «connecteurs» et d'«incises» serait le mot *peut-être*. Sans complémenteur, il interrompt ou suit la proposition comme coda. Toutefois, dans le discours de ces apprenants, il peut introduire des propositions ou de plus grandes unités.

L'intuition qui sous-tend le dénombrement des types lexicaux de ces éléments extra-propositionnels du discours est liée à l'observation que tous les apprenants s'en servent jusqu'à un certain point. Globalement, dans nos corpus, ces éléments sont très nombreux. Si l'on présume qu'ils sont forcément fréquents, on pourrait postuler que la variation lexicale parmi les connecteurs et les incises serait valorisée, en tant qu'indice de la capacité de différencier lexicalement les divers rapports entre unités adjacentes du discours et qu'ainsi, elle pourrait correspondre à des appréciations relativement favorables de ce discours. Nous verrons ci-dessous ce qu'il en est dans le cas de nos informateurs.

À l'aide d'un extrait du corpus d'un de nos apprenants portugais, l'Appendice présente des exemples de formes verbales suivies de leur code numérique et, dans le cas des formes agrammaticales,

d'un astérisque et d'un code correcteur, ainsi que des exemples de connecteurs et d'incises, en contexte. En examinant l'analyse statistique ci-dessous, il faudra retenir qu'un lexème employé et comme connecteur et comme incise est compté deux fois dans l'inventaire des «types». Ceci, manifestement, attribue une valeur supplémentaire à la polyvalence syntaxique des connecteurs-incises.

C. Les évaluations globales

Afin d'établir une comparaison indépendante des deux mesures linguistiques mentionnées ci-dessus, nous avons exploité des évaluations globales de notre corpus, faites par des francophones. Le premier type faisait partie de l'étude des anglophones et des lusophones, dont nous présentons les résultats en détail ci-dessous. Nous caractérisons ce premier type d'évaluation comme «naïf». Il s'agit de le différencier du type appliqué aux entrevues du "Foreign Service Institute" des États-Unis (dorénavant «FSI»), dont il sera question plus bas. C'est que nos évaluateurs «naïfs» n'ont pas reçu de formation ou d'entraînement et leurs évaluations n'ont pas été compartimentées d'avance. Nous avons demandé à six étudiants francophones d'un cours de premier cycle en linguistique appliquée de porter ces jugements individuellement, dans le contexte d'une unité portant sur l'acquisition des L2 et plus particulièrement sur l'analyse et l'évaluation de données linguistiques provenant d'apprenants. Ils savaient qu'il s'agissait d'évaluer la performance linguistique de deux groupes de locuteurs du français L2.

Dans un premier temps, il fallait écouter les entrevues «Balablok» des onze témoins francophones du même âge et du même niveau de scolarité que ceux des vingt-trois informateurs anglophones (voir l'*Introduction*). Nous avons expliqué aux juges qu'il s'agissait d'établir une norme, à laquelle comparer, d'abord, les anglophones qui avaient subi la même entrevue. Ensuite les juges devaient écouter chacun des anglophones et, sans essayer d'analyser le corpus, ils devaient attribuer une note, en pourcentage. Ceci se ferait entièrement en fonction d'une comparaison, globale et intuitive, à la performance des témoins francophones à la même tâche.

Une fois cette première série de jugements terminée, nous avons demandé aux juges d'écouter les entrevues des vingt apprenants portugais et, encore une fois, d'attribuer une note sur cent à chacun d'entre eux, de façon intuitive. Étant donné la nature de l'interview avec les immigrants lusophones, il n'existait pas de groupe témoin francophone ayant subi une entrevue similaire. Nous avons simplement demandé aux juges de se servir de leur expérience auprès des deux groupes précédents de locuteurs en attribuant un score global à chaque apprenant portugais. C'était seulement à la suite de l'audition des trois séries d'entretiens que nous avons demandé aux juges-étudiants d'écrire une description *(post hoc)* des critères qu'ils croyaient avoir employés (toujours de façon intuitive). Cependant, cette étape était uniquement une exigence du cours et ces comptes rendus ne figurent pas dans la présente étude. Nous avons simplement additionné les scores. Cinq juges ont participé à l'évaluation des Portugais et six à celles des anglophones. Les totaux maximums respectifs étaient donc de 500 et de 600.

Les évaluations du type «FSI» sont globales aussi, dans un sens que nous expliquerons ci-dessous, mais elles s'opposent nettement à nos évaluations «naïves» sur certains points importants : premièrement, les locuteurs natifs qui portent les jugements exigés par la méthode «FSI» sont entraînés à adapter leurs impressions à une procédure d'évaluation. Deuxièmement, les «notes» qui constituent le score global correspondent à un compartimentage et à une pondération. Il s'agit, pour le juge, d'attribuer des notes en fonction de jugements portant sur des composantes distinctes de la performance linguistique. C'est-à-dire qu'on évalue «l'accent» sur une échelle de 1 à 6, où les deux extrémités sont étiquetées respectivement «étranger» et «natif»; on évalue la «grammaire» de la même façon mais les bouts de l'échelle sont étiquetés cette fois «incorrecte» et «correcte»; pour le «vocabulaire», il s'agit de «inadéquat» et «adéquat»; pour le «débit», «inégal» vs. «égal», et pour la «compréhension», «incomplète» et «complète». Chaque note distincte (de 1 à 6, correspondant à une position sur une de ces échelles) est multipliée par un coefficient différent, ce qui donne un résultat. La somme de ces derniers peut atteindre un maximum de 99 (Painchaud et coll. 1984 : 72-5). De plus, nous multiplions tous ces scores par dix pour éviter les nombres

décimaux, tel qu'exigé par notre programme statistique, ce qui donne un total maximum de 990.

Dans le cas de l'application spécifique du système d'évaluation «FSI» qui nous intéresse, deux juges francophones du groupe de recherche d'Anglejan-Painchaud se sont entraînés à porter de tels jugements en fonction de ces échelles, à l'aide des enregistrements d'un pré-test du grand échantillon d'immigrants-apprenants dont nous avons étudié un sous-ensemble de dix-sept (voir l'*Introduction*). Grâce à cette préparation sans doute, les coefficients de corrélation (rho) entre l'ensemble des paires d'évaluations par les deux juges, portant sur les corpus d'entrevues dont nous utilisons une partie, étaient de .826 et de .825 respectivement (Ibid : 8, Tab. 1), pour les deux séries d'interviews. La première a eu lieu à la fin d'un cours intensif de français et la seconde, environ sept mois plus tard. Les scores dont nous nous servons sont les moyennes des deux scores (totaux), attribués par les deux juges à notre sous-ensemble des sujets du groupe Painchaud-d'Anglejan pour leur seconde interview (Painchaud et coll. 1984 : 28), puisque nous étudions cette dernière ici.

Étant donné l'entraînement mentionné ci-dessus et la corrélation élevée entre l'ensemble des paires d'évaluations, nous présumons que les évaluations «FSI» sont plus fiables que les cinq ou six évaluations «naïves» mentionnées ci-dessus, que nous avons seulement additionnées et entre lesquelles nous n'avons pas étudié les corrélations. Même si nous utilisons également les rangs engendrés par ces évaluations globales dans les matrices de corrélation Pearson, ci-dessous (du plus fort au plus faible), les scores bruts conservent les données de l'inégalité des écarts entre rangs, et nous les retenons aussi pour cette raison.

Les évaluations «FSI» sont globales en ce sens que les sous-évaluations, de «l'accent», de la «grammaire», du «vocabulaire», du «débit» et de la «compréhension», ne correspondent pas à des analyses linguistiques, mais plutôt à une appréciation intuitive de la composante postulée de la performance linguistique du locuteur. Étant donné qu'il nous faut étudier nos trois ensembles d'informateurs séparément (puisque les tâches imposées par les protocoles d'interview étaient différentes), nous pourrons examiner la relation entre nos évaluations «naïves» et les mesures linguistiques (qui

sont plus analytiques), d'une part, et la relation entre l'évaluation plus systématique, de type «FSI», et ces mêmes mesures linguistiques, d'autre part.

Hypothèses

Plus particulièrement, nous voudrons déterminer les corrélations entre les résultats de chacun des deux types d'évaluation globale et les mesures fondées sur la grammaticalité des formes verbales dans leurs contextes, ou sur la variété des connecteurs et des incises. Nous nous attendrions à une correspondance plus étroite entre les évaluations «naïves» et la variété relative des connecteurs et des incises, en supposant, par exemple, qu'une performance relativement variée au moment d'interrompre le discours pèserait davantage en faveur du candidat lorsque la procédure d'évaluation ne distingue pas les composantes de cette performance linguistique, et ne se prononce surtout pas sur leur correction relative. Nous nous attendrions à une relation moins étroite entre cet indice de la variété des connecteurs-incises et les évaluations «FSI». En revanche, la grammaticalité relative des choix de formes verbales (à l'égard du temps-mode-aspect) devrait correspondre plus étroitement aux évaluations «FSI», même si celles-ci n'impliquent pas de véritable analyse linguistique. C'est que la procédure associée à ce dernier type d'évaluation invoque la grammaticalité, l'adéquation et la correction, c'est-à-dire une comparaison, explicite ou implicite, au sujet parlant «idéal».

Plus précisément, nous prévoyons une corrélation plus élevée entre l'évaluation globale et la grammaticalité verbale qu'entre celle-là et la variété des connecteurs-incises, si l'évaluation globale est de type «FSI». Si, au contraire, l'évaluation globale est «naïve» (et c'est le cas, par exemple, des apprenants portugais, ci-dessous), nous prédisons une corrélation plus élevée de celle-ci, d'une part, avec la variété des connecteurs-incises qu'avec la grammaticalité des formes verbales, d'autre part.

Les résultats chez les lusophones

Pour des raisons qui s'éclairciront ci-dessous, nous traitons des résultats du groupe lusophone en détail ici.

Les Tableaux A et B exposent les résultats pour douze variables, dans le cas des vingt apprenants portugais : le Tableau A indique, pour chaque sujet, la valeur des douze variables, que nous présentons avec leurs abréviations ici :

1. EVA = le score de l'évaluation globale («naïve», dans ce cas - max. 500);
2. RAN = le rang, dérivé de (1), ci-dessus;
3. CON = le nombre de connecteurs distincts (sur le plan lexical);
4. INC = le nombre d'incises lexicalement distinctes;
5. TYP = les «types» c'est-à-dire le nombre de connecteurs et d'incises;
6. VG = le nombre de formes verbales (occurrences) jugées grammaticales à l'égard du temps-mode-aspect;
7. V* = le nombre de formes verbales jugées agrammaticales, au même sens;
8. VT = le nombre total des formes verbales (occurrences);
9. LON = la longueur du corpus en mots orthographiques;
10. VG/VT = verbes grammaticaux/verbes (occurrences);
11. TYP/LON = «types»/longueur du corpus;
12. TYP/LOG LON = «types»/le logarithme naturel de la longueur du corpus.

Le Tableau B présente les coefficients de corrélation (Pearson) entre chaque paire de variables. Les corrélations les plus élevées sont artificielles, dans la mesure où les valeurs de certaines variables sont des sous-produits de la longueur du corpus. Nous verrons plus loin comment corriger l'effet dominateur de cette longueur. Le «rang» entretient une corrélation inverse avec l'évaluation globale, simplement parce que la plus haute évaluation s'est vu attribuer le rang 1 et la plus basse, 18 (il y a deux *ex aequo*), et ainsi de suite.

L'évaluation globale discrimine davantage que le rang, car les rangs ne correspondent pas, en général, à des scores équidistants. La variable «types» n'a pas d'existence indépendante, étant la somme des connecteurs et des incises, mais elle a l'avantage de présenter des valeurs plus élevées que les variables «connecteurs» ou «incises», prises isolément. Il est un peu plus intéressant de noter la corrélation très élevée entre connecteurs et incises dans

La variable «types» n'a pas d'existence indépendante, étant la somme des connecteurs et des incises, mais elle a l'avantage de présenter des valeurs plus élevées que les variables «connecteurs» ou «incises», prises isolément. Il est un peu plus intéressant de noter la corrélation très élevée entre connecteurs et incises dans le cas des lusophones, corrélation due en partie à leurs inventaires, qui se recoupent. (Les connecteurs-incises ont été comptabilisés des deux côtés - voir la *Section B* ci-dessus).

Plus intéressant encore, l'évaluation globale est ici en corrélation presqu'égale avec les «types» (de connecteurs et d'incises), d'un côté, et avec le pourcentage de verbes jugés grammaticaux, de l'autre (.626 et .631, respectivement). Ceci falsifie clairement nos deux hypothèses, pour ce qui est des apprenants portugais (voir *Hypothèses*, ci-dessus). Toutefois, la longueur des corpus est presqu'aussi étroitement liée ici à l'évaluation globale qu'est le nombre de verbes grammaticaux (.715 et .741, respectivement). Ceci résulte du fait que le nombre de verbes (grammaticaux ou non) dépend, en grande partie, de la longueur du corpus. C'est pourquoi cette dernière variable entretient une corrélation de .994 avec le nombre de V grammaticaux et de .996 avec le nombre total de formes verbales. Le pourcentage des verbes jugés grammaticaux (VG/VT) entre dans des rapports plus révélateurs : avec l'évaluation globale (.631), d'une part, et avec les «types» de connecteurs-incises (.419), d'autre part. Cette variable (10, ci-dessus) corrige l'effet de la longueur des corpus.

On ne peut pas en dire autant en ce qui concerne le rapport entre les «types» et la longueur des corpus (variable 11, ci-dessus), pour la simple raison que les valeurs de ceux-là sont noyées dans les valeurs de celleci. Au fur et à mesure que les mots du corpus augmentent (occurrences), les nouveaux types lexicaux (y compris les types de connecteurs et d'incises) y apparaissent de moins en moins souvent.

Avant d'introduire le correctif qui s'impose, cependant, il faut noter les niveaux critiques des corrélations déjà établies (Tab. B), étant donné la directionnalité des hypothèses complémentaires (voir *Hypothèses*) et les dix-huit degrés de liberté pour les vingt sujets étudiés ici.

Le niveau critique des corrélations entre VG/VT, d'un côté et, les connecteurs, les incises, les «types» et la longueur, de l'autre, est 0,05. Dans le cas des corrélations entre l'évaluation globale, d'une part, et, tour à tour, les connecteurs, les incises, les types et VG/VT, d'autre part, $p \leq 0{,}005$. Bien entendu, c'est le cas également des corrélations inverses correspondantes, entre le «rang» (selon l'évaluation globale), d'un côté, et chacun de ces indices, de l'autre. En ce qui a trait aux corrélations entre le nombre de formes verbales grammaticales, d'une part, et l'évaluation globale (ainsi que le «rang»), les connecteurs, les incises et les types, d'autre part, $p \leq 0{,}0005$. C'est le cas aussi de la longueur (et, bien sûr, du nombre total de verbes), d'un côté, et l'évaluation globale (ainsi que le «rang»), de l'autre (Hatch et Farhady 1982 : 277).

Le niveau élevé des autres corrélations significatives est tributaire de la composition des variables : certaines sont les composantes d'autres : les connecteurs et les incises constituent les «types»; les formes verbales G et * constituent le VT. Certaines variables sont déterminées par d'autres : le «rang» par l'évaluation globale, la variable VG/VT par les formes verbales G et *. Le lien très étroit entre le «Total V» et la «longueur» découle de la nature même de la syntaxe. Nous avons vu que les ensembles de «connecteurs» et d'«incises» renferment un sous-ensemble identique, qui est compté dans les deux inventaires, ce qui aide à engendrer une corrélation particulièrement élevée entre les deux, même si l'on pourrait soupçonner que ce n'est pas la seule raison de ce lien étroit. Nous verrons que le nombre de connecteurs et le nombre d'incises ne sont pas toujours étroitement liés.

Les corrélations, très élevées ici, entre le nombre absolu de connecteurs, d'incises (et donc de «types»), de verbes grammaticaux et de verbes, tout court, sont déterminées en plus ou moins grande partie par la longueur des corpus. Il est donc plus intéressant de constater que la corrélation entre le pourcentage de V grammaticaux et le total des V et celle entre VG/VT et la longueur du corpus atteignent à peine le seuil critique. Ainsi, nous voyons que le volume de la production des formes pertinentes n'est pas aussi étroitement lié à la grammaticalité relative de ces formes que l'on aurait cru.

De la même façon, le nombre absolu de formes verbales agrammaticales ne s'associe à aucun autre indice ici, mis à part VG/VT, indice qu'il détermine en partie. Ce résultat surprend seulement dans le cas de l'évaluation globale (et donc le rang) - indices qui s'associent à «V*» dans la direction attendue, sans atteindre le seuil critique. Il semblerait que le nombre absolu d'erreurs de temps (mode, aspect) dans un corpus n'ait pas grand-chose à voir avec les appréciations naïves de la performance linguistique du locuteur.

L'indice corrigé qui reflète le rapport entre les «types» et la «longueur» apparaît finalement ici comme la dernière variable (12) : le nombre de «types» divisé par le logarithme naturel du nombre de mots dans le corpus. En particulier, le remplacement, dans la comparaison, de la longueur par son logarithme corrige l'effet du taux décroissant de l'ajout de n'importe quels types lexicaux, dans n'importe quel type de texte (D. Sankoff, communication personnelle). Ce dernier indice, en définitive, présente un réseau de corrélations relativement symétrique à celui de son rival, VG/VT. Il entretient même une modeste corrélation significative avec celui-ci (au niveau 0,05), une corrélation avec l'évaluation globale et le «rang» au niveau 0,005 et enfin avec la longueur et ses indices tributaires, au niveau 0,0005.

Il semblerait, en résumé, que l'étude du premier groupe d'apprenants ait falsifié nos hypothèses, puisque la grammaticalité relative des choix du temps (mode, aspect), d'un côté, et la variété relative des connecteurs et des incises, de l'autre, ont un rapport très semblable avec les évaluations globales de type «naïf». Pour l'instant, par conséquent, nous n'avons pas la preuve, du moins en ce qui a trait aux juges «naïfs», que la variation lexicale à l'occasion des interruptions discursives est plus cruciale que l'usage correct du système verbal. Nous verrons que cette falsification de nos hypothèses, en faveur des deux indices proposés, aura des conséquences importantes ci-dessous.

Résultats divergents pour les autres groupes

En examinant les données statistiques correspondantes pour les apprenants anglophones (Tab. A' et B', ci-dessous), cependant, il

semble que ni l'un ni l'autre des principaux indices linguistiques n'ait joué le rôle crucial attendu dans l'évaluation globale et «naïve» du groupe par les juges francophones. Le pourcentage de formes verbales jugées grammaticales pour ce qui est du temps-aspect-mode (VG/VT) est loin d'atteindre une corrélation significative avec l'évaluation globale. Le rapport entre les «types» et le logarithme naturel de la longueur du corpus (X12) entretient, de justesse, une corrélation à l'évaluation au niveau 0,05. C'est également le cas de deux indices étroitement apparentés : le nombre absolu de «connecteurs» et de «types» (= «connecteurs» + «incises»). Dans cette mesure, notre hypothèse est confirmée pour les apprenants anglophones, malgré sa falsification chez les lusophones. C'est-à-dire que les indices fondés principalement sur la variété lexicale des connecteurs discursifs semblent, pour cette série de jugements naïfs, plus étroitement liés à l'évaluation globale qu'est la correction relative de l'usage du temps-aspect-mode chez les apprenants.

Dans le cadre de notre étude, il est plus important ici de déterminer pourquoi le pourcentage de verbes grammaticaux, dans le cas des anglophones, n'entretient pas une corrélation aussi élevée avec l'évaluation globale que dans le cas des lusophones. Une raison apparente serait le fait que le choix du temps-aspect-mode ne discrimine pas si nettement parmi ceux-là que parmi ceux-ci (voir, par ex. : aux Tab. A et A', les valeurs de la variable VG/VT, ci-après). Certaines de nos données, que nous ne pourrions passer en revue ici, indiquent qu'une mesure de la grammaticalité qui ne tient compte que du temps-aspect-mode ne reflète guère l'ensemble des difficultés que certains groupes linguistiques éprouvent en utilisant le système verbal en français. Nous songeons à des problèmes typiques des anglophones ou des allophones qui sont beaucoup plus rares chez les lusophones.

Cependant, le fait que X12 - le rapport entre «types» et le logarithme naturel de la longueur du corpus - manifeste une corrélation beaucoup plus élevée avec l'évaluation globale et «naïve» chez les lusophones que chez les anglophones laisse entrevoir un problème supplémentaire : il est possible que les juges francophones n'aient pas employé les mêmes critères en évaluant les Portugais et les anglophones. Il se peut que le contexte sociopolitique, le contenu de l'entrevue et la personnalité (perçue) du

répondant aient joué un rôle plus important dans la détermination de l'évaluation globale («naïve») pour les anglophones que pour les lusophones.

Le protocole d'interview même constituait, peut-être, un déterminant très important des indices linguistiques chez les anglophones. Les matrices de corrélation (Tab. B et B') indiquent que la longueur du corpus s'apparente plus à la plupart des indices chez les lusophones que chez les anglophones. De fait, nous avons prétendu, ci-dessus, que c'est le déterminant essentiel de certains autres indices chez les Portugais. Dans le cas des anglophones (et, nous le verrons, des allophones), ceci est moins vrai. On pourrait supposer que les exigences d'un protocole plus uniforme, notamment chez les anglophones, seraient des déterminants importants de la longueur de l'entrevue et, ainsi, du corpus. La capacité de parler français aurait donc joué un rôle plus important dans la détermination de la longueur du corpus chez les lusophones. Les autres indices linguistiques s'apparenteraient à la longueur parce qu'ils suivent de plus près le profil du langage spontané. Là où la capacité de parler français a moins déterminé la longueur du corpus (c'est-à-dire, dans le cas des apprenants non lusophones), les indices de la performance linguistique qui sont déterminés par la capacité de parler français seraient moins étroitement liés à la longueur du corpus.

Chose surprenante, toutefois, c'est le contraire qui se produit dans le cas de la variable V*, le nombre absolu de formes verbales jugées agrammaticales. Il semblerait que le nombre de ces fautes s'apparente bien à la longueur chez les anglophones et encore plus chez les allophones parce que leur corpus consistaient essentiellement en des réponses à des questions qui exigeaient l'exploitation du système verbal. Ceci était moins vrai des corpus portugais. De toute façon, le nombre absolu de ces erreurs (V*) ne semble pas avoir influencé l'évaluation globale de quel groupe que ce soit (voir les Tab. B, B' et B'', ci-après).

Si la longueur du corpus est moins déterminée par la capacité de parler français chez les anglophones que chez les lusophones, on voit pourquoi les indices qui étaient en corrélation étroite avec la longueur chez ceux-ci le sont moins chez ceux-là (voir les Tab. B et B'). Les grands écarts, dans ce sens, impliquent l'évaluation

globale (et donc le «rang»), les connecteurs, les incises (et ce faisant les «types»), le pourcentage de verbes grammaticaux (VG/VT), et le rapport «types/logarithme de la longueur» (X12). Notamment, le nombre de connecteurs et le nombre d'incises ne se comportent pas de façon analogue, pour ce qui est du rapport à la longueur et aux autres indices connexes chez les lusophones et chez les anglophones. Ce qui est plus important (et non tributaire de la longueur), le nombre de connecteurs n'est pas en corrélation significative avec nombre d'incises, chez les anglophones, à la différence des autres groupes (voir les Tab. B, B' et B''). Nous ne sommes pas en mesure d'expliquer ce résultat.

Les apprenants allophones confirment l'hypothèse pertinente à leurs évaluations. C'est-à-dire que l'évaluation globale («non-naïve») est en corrélation significative avec le pourcentage de formes verbales grammaticales (VG/VT), mais non avec la mesure de la variété des connecteurs et des incises (X12). En effet, les allophones semblent appuyer, d'une manière plus convaincante que les anglophones, l'hypothèse sous-jacente selon laquelle des types divergents d'évaluation globale sont influencés différemment par les facteurs grammaticaux et discursifs.

Si l'on examine les propriétés statistiques des variables qui nous intéressent : à propos des «types» (X5), par exemple, en comparant les anglophones et les allophones, on ne peut prétendre que les deux groupes se distinguent nettement. Les minima, les maxima, les étendues, les moyennes et les coefficients de corrélation avec la longueur du corpus sont presqu'identiques. De façon analogue, X10 (VG/VT) présente des caractéristiques semblables dans les deux groupes et ne permet pas de discriminer très nettement à l'intérieur de ceux-ci.

Malgré ces contraintes et les niveaux modestes des corrélations cruciales (voir les Tab. B' et B''), les anglophones et les allophones présentent le contraste postulé, dans les directions attendues. C'est-àdire que dans le cas des anglophones, les résultats de l'évaluation «naïve» correspondent davantage à la variété lexicale des connecteurs et des incises qu'à la grammaticalité relative des formes verbales. Chez les allophones, le contraste est beaucoup plus net : la variété lexicale des connecteurs et des incises ne manifeste aucun rapport à leur évaluation globale «non-naïve»

(c'est-à-dire à orientation linguistique), tandis que le pourcentage de verbes grammaticaux présente une corrélation significative avec cette évaluation.

Conclusion

Pourquoi les lusophones infirment-ils nos hypothèses complémentaires (ci-dessus), alors que les anglophones et les allophones semblent les confirmer? Nous proposerions une explication en fonction de la nature des groupes et des entrevues. Les lusophones ne constituent un groupe présélectionné ni au point de vue de la formation en français (à la différence des anglophones) ni au point de vue du niveau en français (à la différence des allophones). Même si nous avions l'impression que chacun des trois groupes était hétérogène pour ce qui est de la qualité du français, il reste que les anglophones et les allophones étaient diplômés de programmes d'apprentissage intensif du français - programmes spécifiques à chaque groupe - alors que les lusophones ne partageaient pas ce trait.

Les types d'entrevue se distinguaient entre eux également : l'entretien avec les Portugais était moins constamment «directif» que celui des deux autres groupes. Ainsi, les corpus lusophones sont plus hétérogènes que les deux autres ensembles en ce qui a trait à la longueur (LON, Tab. A, A' et A''). Nous avons déjà vu que la longueur du corpus est en corrélation significative avec l'évaluation globale chez les lusophones, par opposition aux autres groupes (Tab. B, B' et B''), sûrement parce que les lusophones étaient plus libres de parler, ou de se taire, que les autres groupes. Il est possible que le corpus lusophone se rapproche justement, beaucoup plus que les autres, du «langage spontané». Dans cette mesure, les données des Portugais indiqueraient, plus clairement que celles des autres groupes, les facteurs qui interviennent dans l'évaluation «naïve» du discours spontané des apprenants par les locuteurs natifs, et ainsi les indices linguistiques qui se justifient indépendamment comme mesures d'évaluation.

Il est certain que le contraste le plus frappant entre la matrice des corrélations chez les Portugais, d'un côté, et celles des autres

groupes, de l'autre, se trouve dans les niveaux des corrélations entre l'évaluation globale et les indices linguistiques (abstraction faite de la variable VG/VT chez les allophones). Il est possible que ce contraste signifie que le discours spontané, à la différence d'entrevues dirigées, permet l'utilisation, au choix, d'une variété d'indices linguistiques comme mesures du niveau d'acquisition.

Dans cette mesure, les résultats de notre étude représentent des différences importantes entre le discours spontané et «élicité», à l'égard des paramètres qui nous intéressent. Nous proposerions, notamment, qu'en discours spontané, la longueur du corpus de l'apprenant serait déjà un indice de son niveau d'acquisition de la L2 et qu'elle manifesterait une corrélation significative avec d'autres indices d'acquisition qui sont justifiés indépendamment. Dans notre étude, l'évaluation globale et le pourcentage de verbes grammaticaux (VG/VT) semblent particulièrement sensibles à la différence entre la loquacité qui est déterminée par le locuteur et celle qui est déterminée par des facteurs externes. La longueur du corpus est en corrélation significative avec ces deux autres variables dans le cas du discours relativement spontané des lusophones, à la différence du discours moins spontané des autres groupes.

Il était peut-être hâtif de présumer, comme nous l'avons fait en présentant les données portugaises ci-dessus, que certaines de nos variables présentent des corrélations très élevées avec la longueur du corpus pour des raisons purement tributaires de notre façon de procéder. Les différences entre les résultats des trois sous-études, pour ce qui est des coefficients de corrélation entre longueur, d'une part, et le nombre absolu de types de connecteurs, d'incises, ou des deux, d'autre part, indiquent, au contraire, que les corrélations extrêmement élevées chez les Portugais ne sont pas entièrement tributaires de la loquacité : la capacité de produire un grand nombre de connecteurs et d'incises lexicalement distincts reflète, en partie, le niveau d'acquisition. Ce n'est pas uniquement une conséquence nécessaire du fait qu'on continue de parler.

Quant au rapport entre la longueur du corpus et l'évaluation globale par les francophones, on devrait peut-être le considérer comme une conséquence de la perception de la volonté relative

de parler, là où cette volonté est pertinente. Dans le format question-réponse, caractéristique des entrevues auprès des anglophones et des allophones, la volonté relative de parler était moins pertinente que dans les cas des lusophones : certains de ces derniers se sont lancés dans des monologues, d'autres ne l'ont pas fait. Indirectement, la longueur du corpus correspond à l'évaluation globale à cause des capacités linguistiques qui sous-tendent la loquacité. Ainsi, ce n'est sûrement pas par hasard que le pourcentage de verbes grammaticaux corresponde étroitement à l'évaluation globale chez les Portugais : paradoxalement, le protocole d'interview à l'intention des anglophones avait visé la création de contextes qui obligeraient le sujet à exploiter le système verbal (entre autres), et qui fourniraient donc la matière pour une évaluation de la connaissance relative de ce système. Cette «élicitation» de données n'a pas fonctionné comme prévu, dans la mesure où les sujets n'ont pas nécessairement essayé de produire les formes verbales attendues.

Les Portugais, au contraire, étaient moins contraints, parce que leur protocole était moins contraignant et parce qu'ils n'ont pas nécessairement répondu aux questions selon les modèles attendus. Et pourtant, la spontanéité relative de leur discours semble avoir été plus révélatrice que les questions visant l'élicitation de formes cruciales.

La morale à en tirer? Si nous voulons obtenir, en entrevue, des données suffisantes pour la détermination des niveaux d'acquisition des apprenants d'une L2, il faut, paradoxalement, encourager le monologue, et donner la parole à ceux qui en relèvent le défi. Dans ces conditions, le choix de l'un ou de l'autre des indices linguistiques examinés ici perdra son caractère problématique, puisque les différents indices qui se justifient indépendamment mèneront à des résultats analogues.

Références

Clahsen H., J.M. Meisel et M. Pienemann. (1983). *Deutsch als Zweitsprache : der Spracherwerb ausländischer Arbeiter*. Tübingen : Narr.

Connors K. (1983). "Performance measures in L2 : classification and correlations". *Bulletin de l'ACLA*, Vol. 5, n° 2. Montréal : Université de Montréal. Pp. 117-41.

Connors K. (1987). "La loquacité, la variété et la grammaticalité dans l'acquisition d'une langue seconde". *Revue québécoise de linguistique*, Vol. 16, n° 2, Montréal : UQAM. Pp. 11-28.

Connors K. et L. Nuckle. (1986). "The morphosyntax of French personal pronouns and the acquisition/learning dichotomy". *Studies in Romance Linguistics* 24. Dordrecht : Foris. Pp. 225-42.

Connors K., J. Reighard et M.-A. Vieira. (1985). "Evaluation of morphosyntactic development in French as L2". *Bulletin de l'ACLA*, Vol. 7, n° 2, Québec : Université Laval. Pp. 55-64.

Connors K., M.-A. Vieira and J. Reighard. (1986). "The acquisition of French morphosyntax by Montreal lusophones". *The Twelfth LACUS Forum*. Lake Bluff, Illinois : LACUS. Pp. 373-84.

Connors K., M.-A. Vieira, J. Reighard et M. Dillinger. (1987). "Le français des lusophones à Montréal : particularités morphosyntaxiques des variétés discursives". *Actes du XVIII^e Congrès international de linguistique et philologie romanes*, Vol. 4, Ed., D. Kremer. Tübingen : Niemeyer.

Hatch E. et H. Farhady. (1982). *Research design and statistics for applied linguistics*. Rowley, MA : Newbury.

Klein W. et N. Dittmar. (1979). *Developing grammars : the acquisition of German syntax by foreign workers*. Berlin, Heidelberg, N.Y. : Springer.

Painchaud G., A. D'Anglejan et C. Renaud. (1984). *Acquisition du français par des immigrants adultes au Québec*. Québec : CIRB.

Trévise A. et C. Noyau. (1984). "Adult Spanish speakers and the acquisition of French negation forms : individual variation and linguistic awareness", dans Andersen, R. (éd.), *Second languages : a cross-linguistic perspective*. Rowley, MA : Newbury. Pp. 165-89.

Véronique D. (1984). "The acquisition and use of aspects of French morphosyntax by native speakers of Arabic dialects (North Africa)", dans Andersen, R. (éd.). Pp. 91-213.

Annexes et appendice

TABLEAU A Lusophones

SUJET	EVA	RAN	CON	INC	TYP	VG	V*
01	345	17	8	2	10	115	93
02	378	13	6	3	9	118	47
03	350	16	6	6	12	73	21
04	345	17	6	5	11	174	75
05	435	2	15	12	27	573	77
06	423	4	11	10	21	391	33
07	388	11	10	6	16	332	35
08	430	3	11	4	15	315	9
09	385	12	14	7	21	289	18
10	328	18	5	3	8	91	33
11	355	14	10	7	17	301	13
12	401	8	11	10	21	222	18
13	390	10	11	3	14	335	15
14	400	9	11	6	17	403	21
15	400	9	22	13	35	521	21
16	353	15	4	3	7	129	1
17	415	6	6	2	8	243	1
18	443	1	27	23	50	1160	34
19	410	7	13	13	26	517	19
20	420	5	28	27	55	739	16

SUJET	VT	LON	VG/VT	TYP/LON	TYP/ LOG LON
01	208	1112	.553	8.993E-3	1.426
02	165	925	.715	.010	1.318
03	94	714	.777	.017	1.826
04	249	1486	.699	7.402E-3	1.506
05	650	3368	.882	8.017E-3	3.324
06	424	2206	.922	.010	2.728
07	367	2003	.905	7.988E-3	2.105
08	324	1908	.972	7.862E-3	1.986
09	307	1682	.941	.012	2.827
10	124	726	.734	.011	1.214
11	314	1630	.959	.010	2.298
12	240	1314	.925	.016	2.924
13	350	1877	.957	7.459E-3	1.857
14	424	2383	.950	7.134E-3	2.186
15	542	3029	.961	.012	4.366
16	130	751	.992	9.321E-3	1.057
17	244	1299	.996	6.159E-3	1.116
18	1194	6348	.972	7.876E-3	5.710
19	536	3134	.965	8.296E-3	3.230
20	755	4434	.979	.012	6.550

TABLEAU A' Anglophones

SUJET	EVA	RAN	CON	INC	TYP	VG	V°
01	399	19	5	3	8	208	30
02	465	4	11	2	13	240	44
03	445	7	9	5	14	322	20
04	415	15	5	4	9	224	18
05	441	10	8	5	13	194	20
06	385	20	6	0	6	124	16
07	475	3	9	5	14	187	24
08	495	1	10	8	18	294	26
09	462	5	15	3	18	264	17
10	412	16	9	3	12	272	48
11	435	12	11	11	22	370	41
12	447	6	11	2	13	205	6
13	382	21	6	3	9	175	31
14	465	4	9	3	12	247	21
15	372	22	9	6	15	258	36
16	443	8	9	3	12	224	60
17	400	18	10	11	21	255	31
18	405	17	16	3	19	141	29
19	425	14	6	1	7	129	7
20	477	2	16	5	21	235	24
21	440	11	7	6	13	131	31
22	432	13	13	5	18	382	69
23	442	9	12	4	16	263	43

TABLEAU A' Anglophones (suite)

SUJET	VT	LON	VG/VT	TYP/LON	TYP/ LOG LON
01	238	1386	.874	5.772E-3	1.106
02	284	1978	.845	6.572E-3	1.713
03	342	2077	.942	6.740E-3	1.833
04	242	1197	.926	7.519E-3	1.270
05	214	1304	.907	.010	1.812
06	140	817	.886	7.344E-3	.895
07	211	1142	.886	.012	1.988
08	320	1754	.919	.010	2.410
09	281	1423	.940	.013	2.479
10	320	1765	.850	6.799E-3	1.605
11	411	2180	.900	.010	2.862
12	211	1067	.972	.012	1.864
13	206	1158	.850	7.772E-3	1.276
14	268	1632	.922	7.353E-3	1.622
15	294	1710	.878	8.772E-3	2.015
16	284	1444	.789	8.310E-3	1.649
17	286	1689	.892	.012	2.826
18	170	1099	.829	.017	2.713
19	136	862	.949	8.121E-3	1.036
20	259	1475	.907	.014	2.878
21	162	953	.809	.014	1.895
22	451	2402	.847	7.494E-3	2.312
23	306	1700	.859	9.412E-3	2.151

TABLEAU A'' ALLOPHONES

SUJET	EVA	RAN	CON	INC	TYP	VG	V*
01	480	9	6	4	10	144	21
02	540	5	6	3	9	168	36
03	580	3	6	6	12	398	49
04	545	4	5	3	8	160	12
05	500	8	5	2	7	166	19
06	625	1	14	5	19	340	38
07	515	7	8	2	10	240	48
08	460	10	8	4	12	364	74
09	535	6	10	6	16	174	16
10	580	3	9	6	15	229	14
11	580	3	9	7	16	160	11
12	455	11	14	9	23	276	61
13	590	2	12	10	21	411	46
14	345	14	3	3	6	57	17
15	430	12	10	5	15	124	12
16	430	12	7	4	11	97	22
17	350	13	13	9	22	350	54

SUJET	VT	LON	VG/VT	TYP/LON	TYP/ LOG LON
01	165	1085	.873	9.217E-3	1.431
02	204	1785	.824	5.042E-3	1.202
03	447	2911	.890	4.122E-3	1.504
04	172	1460	.930	5.479E-3	1.098
05	185	1332	.897	5.255E-3	.973
06	378	2326	.899	8.169E-3	2.451
07	288	2090	.833	4.785E-3	1.308
08	438	3299	.831	3.637E-3	1.481
09	190	1487	.916	.011	2.190
10	243	1746	.942	8.591E-3	2.009
11	171	1274	.936	.013	2.238
12	337	2271	.819	.01	2.976
13	457	2509	.899	8.370E-3	2.683
14	74	675	.770	8.889E-3	.921
15	136	1016	.912	.015	2.166
16	119	1356	.815	8.112E-3	1.525
17	404	3483	.866	6.316E-3	2.698

TABLEAU B LUSOPHONES

	MATRICE DE CORRÉLATION POUR LES VARIABLES X1 À X12											
	EVA	RAN	CON	INC	TYP	VG	V°	VT	LON	VG/VT	TYP/LON	TYP/LO.
EVA	1											
RAN	-.989 D	1										
CON	.645 C	-.637 C	1									
INC	.58 C	-.594 C	.918 D	1								
TYP	.626 C	-.629 C	.979 D	.979 D	1							
VG	.741 D	-.752 D	.9 D	.863 D	.901 D	1						
V°	-.2	.205	-.078	-.067	-.074	-.056	1					
VT	.722 D	-.733 D	.893 D	.858 D	.894 D	.995 D	.041	1				
LON	.715 D	-.723 D	.909 D	.882 D	.915 D	.994 D	.018	.996 D	1			
VG/VT	.631 C	-.639 C	.433 A	.387 A	.419 A	.498 B	-.786 D	.423 A	.426 A	1		
TYP/LON	-.259	.257	.069	.185	.13	-.215	-.186	-.233	-.207	-.103	1	
TYP/LO...	.611 C	-.613 C	.976 D	.978 D	.998 D	.875 D	-.087	.867 D	.889 D	.418 A	.188	1

A ---> p < ou = .05
B ---> p < ou = .025
C ---> p < ou = .005
D ---> p < ou = .0005

TABLEAU B' ANGLOPHONES

	MATRICE DE CORRÉLATION POUR LES VARIABLES X1 À X12											
	EVA	RAN	CON	INC	TYP	VG	V°	VT	LON	VG/VT	TYP/LON	TYP/LO.
EVA	1											
RAN	-.985 E	1										
CON	.407 A	-.423 B	1									
INC	.11	-.045	.167	1								
TYP	.353 A	-.326	.807 E	.717 E	1							
VG	.258	-.239	.356 A	.538 D	.574 D	1						
V°	-.075	.113	.21	.224	.283	.506 C	1					
VT	.214	-.19	.357 A	.52 C	.564 D	.986 E	.645 D	1				
LON	.212	-.194	.357 A	.491 C	.546 D	.94 E	.63 D	.957 E	1			
VG/VT	.252	-.282	-.017	.011	-5.17E-3	.127	-.756 E	-.036	-.066	1		
TYP/LON	.224	-.214	.615 D	.279	.602 D	-.257	-.228	-.273	-.311	-.022	1	
TYP/LO...	.354 A	-.329	.82 E	.691 E	.994 E	.486 C	.224	.475 B	.454 B	-6.67E-3	.686 E	1

A ---> p < ou = .05
B ---> p < ou = .025
C ---> p < ou = .01
D ---> p < ou = .005
E ---> p < ou = .0005

TABLEAU B'' ALLOPHONES

	MATRICE DE CORRÉLATION POUR LES VARIABLES X1 À X12											
	EVA	RAN	CON	INC	TYP	VG	V°	VT	LON	VG/VT	TYP/LON	TYP/LO.
EVA	1											
RAN	-.977 E	1										
CON	.154	-.135	1									
INC	.075	-.118	.751 E	1								
TYP	.118	-.124	.958 E	.908 E	1							
VG	.347	-.386	.559 C	.555 B	.585 C	1						
V°	-.094	.098	.406	.315	.39	.776 E	1					
VT	.288	-.322	.551 B	.534 B	.572 C	.995 E	.834 E	1				
LON	.081	-.132	.499 B	.462 A	.515 B	.915 E	.869 E	.936 E	1			
VG/VT	.628 D	-.641 D	.222	.242	.244	.151	-.43 A	.064	-.045	1		
TYP/LON	-.083	.138	.336	.353	.371	-.41	-.52 B	-.439 A	-.536 B	.283	1	
TYP/LO...	.118	-.113	.948 E	.899 E	.991 E	.489 B	.286	.472 A	.402	.28	.491 B	1

A ---> p < ou = .05
B ---> p < ou = .025
C ---> p < ou = .01
D ---> p < ou = .005
E ---> p < ou = .0005

1..1 $ Quand vous êtes venu au Canada quels étaient les
1..2 premiers problèmes que vous avez eus? %
1..2
1..3 je pensa1 que - quand nous - quand nous
1..4 arrivons1*3 - nous ne parlons1*2 presque rien de
1..5 français - c'est1*2 - [[alors]] c'est1*2 difficile
1..6 de comprendre9 - [[mais]] avec le temps - nous -
1..7 nous allons le comprendre0 beaucoup mieux -
1..8 [[mais]] même maintenant c'est1 trop difficile
1..9 encore de parler9 et de m'exprimer9 en français -
1.10 parce que je n'ai pas suivi3 aucune cours de
1.11 français - excepté un cours dans le centre
1.12 portugais de la [[mais]] ce cours-là - c'est1*2
1.13 pour prendre9 - débutants - c'est1 ça? - [[/i/]] -
1.14 [[/i/ alors]] - tous ce que - il apprenait2 là je
1.15 savais2 déjà - parce que j'ai1*3 un cours de
1.16 français d'aujourd''hui le - c'est1 le sillon de
1.17 que j'ai apporté3 avec moi de Portugal - [[mais]]
1.18 le travail - le - <<bon>> - la vie - c'est1
1.19 quotidien c'est1 difficile de - de mélanger9 tout
1.20 ça
1.20
1.21 $ Est-ce que vous parlez le français avec des amis
1.22 ou au travail? %
1.22
1.23 non - au travail je - nous sommes1 un groupe - de
1.24 neuf personnes - toutes Portugais - au travail je
1.25 parle1 portugais seulement - [[excepté]] des fois
1.26 nous parlons1 français avec des gens qui nous
1.27 demandons1 quelque question - [[excepté]] - nous
1.28 parlons1 portugais seulement - pour moi je pense1
1.29 que c'est1 très difficile - plus difficile ça
1.30 parce que nous parlons1 seulement portugais - chez
1.31 nous - nous parlons1 portugais aussi
1.31
1.32 $ Mais vos enfants parlent bien le français? %
1.32
1.33 ma fille parle1 portugais - mais français
1.34 <<peut--être>> encore mieux
1.34
1.35 $ Pensez vous qu'ils vont garder la connaissance du
1.36 portugais? %
1.36
1.37 oui - elle va1 déjà à l'école portugaise tous les
1.38 samedis - [[et]] - il ne va1 pas encore - [[mais]]
1.39 - quand il ferait8*6 six ans sept ans il va1*6
1.40 encore là--bas - pour - c'est1 mieux - de - de
1.41 nous a - <<comme dire9>> - pour - pour - pour eux
1.42 de - pour

<< >> = incise
[[ ]] = connecteur

Synthèse

Kathleen Connors

La recherche dont nous avons présenté certains résultats ici avait pour objectif immédiat la description de certains aspects du français d'apprenants adultes. Les critères implicites du choix des aspects à étudier étaient l'intérêt inhérent des structures (sur le plan linguistique) et la possibilité qu'elles soient des indices d'acquisition du français L2. Pour ce qui est de l'intérêt linguistique des structures étudiées, ce sera surtout au lecteur à en juger. Nous pouvons seulement lui rappeler le rôle fondamental du système verbal, de la pronominalisation et de tous les procédés cohésifs dans la communication. Même si les constructions verbales pronominales semblent moins importantes, vues de cet angle, leur complexité morpho-syntaxique et leur statut problématique sur le plan lexico-sémantique leur réservent une place privilégiée parmi les structures typiquement françaises pour lesquelles on pourrait prévoir des problèmes d'acquisition.

La question de savoir si tel ou tel aspect du français des apprenants se prête au statut d'indice d'acquisition du français L2 devrait, au contraire, se laisser poser au niveau empirique. Dans l'*Introduction*, nous avons proposé trois critères d'indices dans ce sens : que l'indice proposé permette de situer les apprenants à différents stades ou niveaux d'acquisition, qu'il soit associé à des formes linguistiques obligatoires (dans certains contextes généraux) ou fréquentes pour d'autres raisons, dans les productions «natives», et qu'il soit associé à des formes linguistiques qui émergent ou qui se stabilisent dans un ordre invariable. Comme nous l'avons vu, faute d'études longitudinales, nous ne pouvons prétendre avoir satisfait qu'aux deux premiers critères ici.

La recherche des indices d'acquisition, dans ce sens précis, n'était cependant pas notre seul objectif. Nous cherchions aussi des éléments ou des structures dont l'acquisition chez les adultes permettrait de faire ressortir des effets de l'interaction entre la grammaire de la L2 et celle de la L1 des apprenants. Ainsi, par exemple, nous avons choisi les constructions verbales pronominales dans l'espoir que l'analyse de leur utilisation chez les anglophones, par opposition aux lusophones, fasse sentir l'impact différentiel de la présence (ou l'absence) d'une construction analogue en L1. Nous pensons avoir atteint cet objectif dans le cas de cette étude.

Si nous semblons préoccupés par l'acquisition des petits détails du français, c'est qu'il faut trouver la spécificité de ce processus : dans quelle mesure nos informateurs adultes acquièrent-ils une grammaire «nouvelle»? Ce n'est pas, par exemple, l'ordre SVO (de base) qui pourrait constituer une nouvelle acquisition pour la plupart de nos sujets, comme nous l'avons vu dans l'*Introduction*. L'existence d'une sous-catégorie de verbes intrinsèquement pronominaux, par contre, constitue un aspect de la grammaire que le portugais partage avec le français, par opposition à l'anglais. Sur le plan lexical, toutefois, les différences de détail dans la sous-catégorisation des verbes influent sur la nature des erreurs et des sur-généralisations que les lusophones sont susceptibles de faire en français.

En définitive, les grandes catégories linguistiques dont nous avons étudié l'usage ici ne sont pas particulières au français : le temps, le mode, l'aspect, les connecteurs, les incises, les pronoms et les autres éléments cohésifs. C'est le déploiement des différentes réalisations de ces catégories, dans des contextes précis, qui caractérisent les divers sujets parlants. Finalement, c'est le degré auquel les apprenants arrivent à maîtriser la complexité du système d'exploitation de ces catégories, dans ces contextes, qui les distingue des locuteurs natifs et qui nous force à reconnaître l'énormité de leur tâche.

C'est dans ce sens-là qu'il faut essayer de comprendre les résultats des comparaisons que nous avons faites entre groupes d'apprenants, d'une part, et entre apprenants et francophones, d'autre part : dans le premier article de Bazergui, les trois groupes

d'apprenants à l'étude sont comparés entre eux pour ce qui est de l'exploitation globale des oppositions de temps, de mode et d'aspect. Même si, sur certains points, cette comparaison générale semble favoriser les lusophones, on ne peut en conclure qu'ils soient vraiment plus forts en français que les autres groupes, puisque leur supériorité apparente pourrait résulter des différences entre les tâches linguistiques imposées par les protocoles d'interview, qui étaient distincts pour chaque groupe d'apprenants.

C'est seulement en comparant l'usage de la construction verbale pronominale que le second article de Bazergui révèle un écart très net entre la maîtrise du système verbal chez les anglophones et les lusophones. En outre, cette comparaison passe obligatoirement par une confrontation des données des entrevues anglophones et lusophones à celles des groupes témoins francophones qui ont répondu à des protocoles d'interviews comparables à ceux de chaque groupe d'apprenants. C'est dans les résultats de cette deuxième étude que Bazergui expose des différences entre la performance des anglophones et des lusophones qui ne peuvent résulter entièrement de la forme des entretiens. Plus important, la comparaison des constructions verbales pronominales chez ces deux groupes d'apprenants fait ressortir, bien plus clairement qu'ailleurs, la contribution de la L1 au développement de la compétence en L2.

Majkrak, en revanche, en abordant les procédés très généraux du discours, compare les anglophones directement et exclusivement au groupe témoin francophone qui a subi la même entrevue : le fait de consacrer un article en entier à l'exploitation du système pronominal par les deux groupes se justifie, grâce à l'importance primordiale de la pronominalisation en français, surtout comme procédé cohésif. Les observations quantitatives illustrent l'importance de l'écart entre la maîtrise du système pronominal chez les anglophones et les francophones.

Sur un autre plan, l'examen détaillé, dans un deuxième temps, du déploiement d'autres procédés cohésifs (la paraphrase, la récurrence, le parallélisme et l'ellipse) souligne la difficulté qu'éprouvent les anglophones à soutenir un discours cohésif en français - difficulté provenant manifestement d'une maîtrise insuffisante de la L2, car il s'agit de procédés discursifs très

généraux que les mêmes sujets auraient déployés automatiquement en anglais. Encore une fois, la quantification des observations dissipe les doutes éventuels : il ne pourrait s'agir d'une vague impression ou d'une légère difficulté (chez les anglophones) à mener le discours. Les écarts de performance, à l'égard des procédés discursifs, touche trop d'aspects distincts, et ce, dans des proportions relativement importantes.

La dernière étude linguistique, celle de Connors, relève de la comparaison interne des groupes d'apprenants à l'étude (les anglophones, les lusophones et les allophones). On y voit les effets quantitatifs, et plus particulièrement statistiques, de l'hétérogénéité considérable des groupes d'apprenants, qui avaient été constitués en fonction de la L1, et/ou des classements associés à des programmes d'apprentissage du français L2. Le véritable objectif de l'article, cependant, est l'évaluation de différents indices d'acquisition du français L2 : parmi les types de caractéristiques linguistiques des apprenants que l'auteur a examinés, lesquels se prêtent à l'évaluation du niveau d'acquisition? Il appert que divers traits morphosyntaxiques ou discursifs peuvent servir d'indices, pourvu qu'ils soient obligatoires et fréquents. À ceci s'ajoute une condition essentielle : le corpus à l'étude doit relever du langage relativement spontané, par opposition à une suite de questions-réponses.

En définitive, c'est à l'aide de nos corpus publiés au tome II de ce recueil, auquel l'article de Lenoble ici même sert de guide détaillé, que le lecteur pourra vérifier nos différents résultats et nos conclusions générales. À la fin, l'unité des travaux présentés ici se dégage dans la constatation que l'étude du discours en L2 permet des comparaisons évaluatives dans la mesure où l'on arrive à identifier des structures linguistiques qui sont à la fois sujettes à la variation acquisitionnelle et fréquentes dans le discours. Elles peuvent être fréquentes soit parce qu'elles constituent les moyens obligatoires (dans la langue cible) d'exprimer des rapports logiquement nécessaires au type de discours en question, soit parce qu'elles caractérisent la langue fondamentalement, abstraction faite du type de discours.

Bibliographie générale de l'acquisition du français L2 chez les adultes

Kathleen Connors

Cette bibliographie s'adresse à tous ceux qui s'intéressent à l'acquisition du français langue seconde chez les adultes. Quelques précisions s'imposent, cependant. Nous n'y avons inclus que les articles et les recueils qui sont, en principe du moins, disponibles dans les bibliothèques universitaires, en excluant les communications, les thèses, les mémoires et les rapports, s'ils sont inédits. En outre, nous nous sommes limités aux articles (ou aux recueils contenant des articles) qui constituent, au moins en partie, des comptes rendus de recherches descriptives et empiriques. Il s'agit là d'un parti pris en faveur d'une sous-catégorie des travaux pertinents, dont les exemplaires sont beaucoup plus rares que les études purement théoriques ou pédagogiques.

Le mot «adulte», dans ce contexte, comprend les personnes ayant au moins dix-huit ans, ce qui permet d'inclure les études sur le français des apprenants aux divers niveaux post-secondaires des systèmes scolaires ou universitaires de plusieurs pays.

Alber J-L. et B. Py. (1986). "Vers un modèle exolingue de la communication interculturelle : interparole, coopération et conversation", dans *Études de linguistique appliquée* 61, pp. 78-90.

Andersen R.. (éd.). (1984). *Second languages : a cross-linguistic perspective*, Rowley, MA, Newbury.

Anonyme. (éd.). (1983 et 1985). *Acquisition du français par des travailleurs marocains. Papiers de travail* 1 et 2.

Anonyme. (éd.). (1987). *Actes du 2e colloque de linguistique hispanique*, Rouen, Presses de l'Université de Haute-Normandie.

Arditty J. et M. Mittner. (éds). (1980). *Acquisition d'une langue étrangère.* Numéro spécial d'*Encrages*, Paris, Presses universitaires de Vincennes.

Bonnot J.-F. et J.J. Spa. (1988). "De la structure théorique de l'interlangue à l'apprentissage du français langue étrangère", dans *Revue internationale de Linguistique appliquée* 26.1, pp. 19-34.

Buteau M.F. (1970). "Student errors and the learning of French as a second language : a pilot study", dans *Revue internationale de linguistique appliquée* 8.2, pp. 133-45.

Chevalier J.-C. (éd.). (1987). *Hommage à Maurice Molho*, Paris, Cahiers de l'E.N.S. de Fontenay.

Connors K. et L. Jasmin-Demers. (1982). "Étude sur le français des diplômés de l'immersion tardive", dans *Eighth LACUS Forum*, Columbia, S.C., Hornbeam Press, pp. 512-22.

Connors K. (1983). "Performance measures in L2 : classification and correlations", dans *Bulletin de l'ACLA* 5.2, pp. 117-41.

Connors K. et M. Toker. (1984). "Analyses quantitatives et qualitatives des test cloze : syntaxe et sémantique", dans *Revue canadienne des langues vivantes* 40, pp. 245-63.

Connors K., J. Reighard et M.-A. Vieira. (1985). "Evaluation of morphosyntactic development in French as L2", dans *Bulletin de l'ACLA* 7.2, pp. 55-64.

Connors K. et L. Nuckle. (1986). "The morphosyntax of French personal pronouns and the acquisition/learning dichotomy", dans *Studies in Romance Linguistics* 24, Dordrecht, Foris, pp. 225-42.

Connors K., J. Reighard et M.-A. Vieira. (1986). "The acquisition of French morphosyntax by Montreal lusophones", *Twelfth LACUS Forum*, Lakebluff, Illinois, Jupiter Press, pp. 373-84.

Connors K. (1987). "La loquacité, la variété et la grammaticalité dans l'acquisition d'une langue seconde", dans *Revue québécoise de linguistique* 16.2, pp. 11-28.

Connors K. (1988). "Morphological markedness in second language acquisition", dans *Revue internationale de linguistique appliquée* 26.2, pp. 101-13.

Connors K., M.-A. Vieira, J. Reighard et M. Dillinger. (1988). "Le français des lusophones à Montréal : particularités morphosyntaxiques des variétés discursives", dans D. Kremer. (éd.), *Actes du XVIII^e Congrès international de Linguistique et de Philologie romanes*, tome V, Tübingen, Niemeyer, pp. 307-19.

Coppieters R. (1987). "Competence differences between native and near-native speakers", dans *Language* 63.3, pp. 544-73.

Dabène L. (éd.). (1981). *Langues et migrations*, Grenoble, Publications de l'Université des Langues et Lettres de Grenoble.

D'Anglejan A. et C. Renaud. (1981). "Variables psychologiques et sociales reliées à l'apprentissage d'une langue seconde par l'immigrant adulte en situation scolaire", dans J.-D. Gendron et R. Vigneault. (éds), *Les mécanismes psychologiques sous-jacents à l'apprentissage d'une langue seconde*, Québec, CIRB, pp. 1-14.

D'Anglejan A., C. Renaud, R.-H. Arsenault et A.-M. Lortie. (1981). *Difficultés d'apprentissage de la langue seconde chez l'immigrant adulte en situation scolaire : une étude dans le contexte québécois*, Québec, CIRB.

D'Anglejan A. et G. Painchaud. (1984). *Acquisition du français parmi des immigrants adultes au Québec : rapport de recherche*, Québec, CIRB.

D'Anglejan A. et C. Renaud. (1985). "Learner characteristics and second language acquisition : a multivariate study of adult immigrants and some thoughts on methodology", dans *Language Learning* 35, pp. 1-19.

D'Anglejan A., G. Painchaud et C. Renaud. (1986). "Beyond the language classroom : A study of communicative abilities in adult immigrants following intense instruction", dans *TESOL Quarterly* 20.2, pp. 185-205.

Dechert H.W., D. Möhle et M. Raupach. (éds). (1984). *Second language productions*, Tübingen, Narr.

Dommergues Y. et B. Grandcolas. (éds). (1980). *Acquisition d'une langue étrangère. Champs éducatifs* I, Paris, Presses universitaires de Vincennes.

European Science Foundation. (1988). *Second language acquisition by adult immigrants*, 6 tomes, Strasbourg, Paris, Heidelberg et Londres, ESF.

Extra G. et M. Mittner. (éds). (1984). *Studies in second language acquisition by adult immigrants*, Tilburg. (Pays-Bas), Université de Tilburg.

Favreau M. et N.S. Segalowitz. (1982). "Second language reading in fluent bilinguals", dans *Applied Psycholinguistics* 3.4, pp. 329-41.

Gardner R.C., R.N. Gardner, R. Moorcroft et F.T. Evers. (1987). "Second language attrition : the role of motivation and use", dans *Journal of Language and Social Psychology* 6.1, pp. 29-47.

Giacomi A. et D. Véronique. (éds). (1986). *Acquisition d'une langue étrangère. Perspectives et recherches*, Aix-en-Provence, Service de Publications de l'Université de Provence.

Greidanus T. et E. Van der Linden. (1986). "Relations between FL grammaticality judgments and FL production", dans *ITL : A Review of Applied Linguistics* 73, pp. 51-82.

Hamm C. (1988). "The French language speaking proficiency of students entering first-year university", dans *Bulletin de l'ACLA* 10.1, pp. 25-34.

Kraushaar B. et S. Lambert. (1987). "Shadowing proficiency according to ear of input and type of bilinguality", dans *Bulletin de l'ACLA* 9.1, pp. 17-31.

Jasmin-Demers L. (1982). "Des tests pragmatiques et multi-fonctionnels pour une évaluation objective de la compétence à communiquer", dans *Bulletin de l'ACLA* 4.2, pp. 115-24.

Lambert W.E. (1972). *Language, psychology, and culture : essays by Wallace E. Lambert*, Stanford, CA, Stanford University Press.

Langages 57. (1980), Paris, Larousse.

Langages 84. (1986), Paris, Larousse.

Langue française 29. (1976), Paris, Larousse.

Langue française 71. (1986), Paris, Larousse.

Leblanc R. et G. Painchaud. (1981). "Auto-évaluation dans l'apprentissage du français langue seconde", dans P.R. Léon et J. Yashinsky. (éds), *Options nouvelles en didactique du français langue étrangère*, Montréal, Didier, pp. 127-36.

Leblanc R., G. Painchaud et G. Monette. (1982). "Évaluation de l'habileté à communiquer et auto-évaluation", dans *Bulletin de l'ACLA* 4.2, pp. 125-35.

Lefranc Y. et K. Sefta. (1982). "Les immigrés de la 'deuxième génération'. Quelle(s). langue(s)? Quelle(s). culture(s)?", *Langage et Société* 22, pp. 3-14.

Lüdi G. (éd.). (1987). *Devenir bilingue, parler bilingue*, Tübingen, Niemeyer.

Lüdi G. et B. Py. (1986). *Être bilingue*, Berne, Lang.

Mallet C.Y. (1987). "Le moniteur chez des apprenants de langue seconde", dans *Bulletin de l'ACLA* 9.1, pp. 33-9.

Manessy G. et P. Wald. (1984). *Le français en Afrique Noire tel qu'on le parle tel qu'on le dit*, Paris, Éditions L'Harmattan.

Meara P. et S. Ingle. (1986). "The formal representation of words in an L2 speaker's lexicon", dans *Second Language Research* 2.2, pp. 160-71.

Ménard N. et K. Connors. (1979). "Recherches sur la compréhension du français québécois par les élèves anglophones", dans *Le français dans le monde* 145, pp. 100-4.

Murray D.J. (1986). "Characteristics of words determining how easily they will be translated into a second language", dans *Applied Psycholinguistics* 7.4, pp. 353-72.

Noyau C. (éd.). (1981). *Les travailleurs étrangers et la langue*, Paris, GRECO 13-CNRS et Presses universitaires de Vincennes.

Noyau C. et R. Porquier. (éds). (1984). *Communiquer dans la langue de l'autre*, Paris, Presses universitaires de Vincennes.

Noyau C. et D. Véronique. (éds). (1986). *Studies in second language acquisition* 8.3.

Olynyk M., D. Sankoff et A. D'Anglejan. (1983). "Second language fluency and the subjective evaluation of officer cadets in a military college", dans *Studies in Second Language Acquisition* 5.2.

Olynyk M., A. D'Anglejan et D. Sankoff. (1987). "A quantitative and qualitative analysis of speech markers in the native and the second language speech of bilinguals", dans *Applied Psycholinguistics* 8.2, pp. 121-36.

Painchaud G. (1979). "Quelques caractéristiques du comportement linguistique des apprenants lents", dans *Travaux de recherche sur le bilinguisme* 18, pp. 1-23.

Painchaud G. et R. Leblanc. (1980). "L'autoévaluation : le cas des langues secondes", dans *Proceedings of the 21st Adult Education Research Conference*, Vancouver, pp. 157-62.

Painchaud G., A. D'Anglejan et C. Renaud. (1984). *Acquisition du français par des immigrants adultes au Québec*, Québec, CIRB.

Painchaud-Leblanc G. et R. Leblanc. (1984). "L'autoévaluation en contexte scolaire", dans *Études de linguistique appliquée* 56, pp. 88-98.

Perdue C. et R. Porquier. (éds). (1979). *Linguistique appliquée*. Numéro spécial d'*Encrages*.

Perdue C. (éd.). (1984). *Second language acquisition by adult immigrants : a field manual. The European Science Foundation Project*. Rowley, MA, Newbury.

Piétropaolo-Saura B. et M. Roffé. (1985). "Analyse différentielle et analyse des erreurs : un nouveau regard", dans *Bulletin de l'ACLA* 7.1, pp. 47-55.

Piquette E. (1977). "L'aptitude du traducteur à déceler l'ambiguïté syntaxique", dans R. Sarrasin. (éd.), *Psycholinguistique expérimentale et théorique*, Montréal, Presses de l'Université du Québec, pp. 169-93.

Porquier R. et A. Bernard. (1988). "Processus d'acquisition et interactions sociales chez des adultes migrants en milieu naturel et chez des adolescents migrants en milieu scolaire", dans *Bulletin CILA* 47, 43-51.

Py B. (éd.). (1984). *Acquisition d'une langue étrangère* III, Paris, Presses universtaires de Vincennes.

Ragusich N-C. (1977). *Contribution à l'étude du problème de la difficulté en langue étrangère*, Québec, CIRB.

Robitaille C. (1987). "Difficultés d'apprentissage en L2 reliées à la structure factitive française : *faire V*", dans *Revue québécoise de linguistique* 16.2, pp. 85-101.

Saunders G.E. (1988). "The structure of errors in the perception of French speech", dans *Revue de phonétique appliquée* 86, pp. 43-99.

Schneiderman E., J.S. Bourdages et C.M. Champagne. (1984). "Evaluating 'accent' in a second language", dans J. Den Haese et J. Nivette. (éds), *AILA Brussels 84 : Proceedings*, Instituut voor Taalonderwijs, Vrije Universiteit Brussel, tome 2, p.830.

Schneiderman E. et coll. (1988). "Second-language accent : the relationship between discrimination and perception in acquisition", dans *Language Learning* 38.1, pp. 1-19.

Singleton D. (1987). "Mother and other tongue influence on learner French", dans *Studies in Second Language Acquisition* 9.3, pp. 327-45.

Tabouret-Keller A. (1975). "Plurilinguisme : Revue des travaux français de 1945 à 1973", dans *La Linguistique* II, pp. 123-37.

Tardif C. et A. D'Anglejan. (1981). "L'influence des erreurs en français langue seconde sur la communication orale", dans *Revue canadienne des langues vivantes* 37.4, pp. 706-24.

Toker M., K. Connors et N. Ménard. (1981). "The effectiveness of 'late immersion' for second language learning", dans *AILA Proceedings* I, Lund, Wallin et Dalholm, pp. 65-6.

Towell R. (1987). "Variability and progress in the language development of advanced learners of a foreign language", dans R. Ellis. (éd.), *Second Language Acquisition in Context*, Englewood Cliffs, N.J., Prentice-Hall.

Towell R. (1987). "Why is variability part of language learning?", dans *TEANGA* 7, pp. 1-34.

Trévise A. (éd.). (1982). *Acquisition d'une langue étrangère* II. Numéro thématique d'*Encrages*, Paris, Presses universitaires de Vincennes.

Whalen K. (1988). "Pilot study on the nature of difficulties in written expression in a second language : process or product?", dans *Bulletin de l'ACLA* 10.1, pp. 51-7.

White L. (1987). "Markedness and second language acquisition : the question of transfer", dans *Studies in Second Language Acquisition* 9.3, pp. 261-86.

White L. (sous presse). "The principle of adjacency in second language acquisition : do L2 learners observe the subset principle?", dans S. Gass et J. Schachter. (éds), *Linguistic Perspectives on Second Language Acquisition*, Cambridge, CUP.

White L. (à paraître). "Argument structure in second language acquisition", dans *Proceedings of the La Trobe University Workshop on Second Language Acquisition*.

Gouvernement du Québec
Office de la langue française

5225
Imprimé au Québec, Canada